This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

HyperSpace: Hypernetworks for spacing-adaptive image segmentation

Samuel Joutard, Maximilian Pietsch, and Raphael Prevost

ImFusion, Munich, Germany
joutard@imfusion.com

Abstract. Medical images are often acquired in different settings, requiring harmonization to adapt to the operating point of algorithms. Specifically, to standardize the physical spacing of imaging voxels in heterogeneous inference settings, images are typically resampled before being processed by deep learning models. However, down-sampling results in loss of information, whereas upsampling introduces redundant information leading to inefficient resource utilization. To overcome these issues, we propose to condition segmentation models on the voxel spacing using hypernetworks. Our approach allows processing images at their native resolutions or at resolutions adjusted to the hardware and time constraints at inference time. Our experiments across multiple datasets demonstrate that our approach achieves competitive performance compared to resolution-specific models, while offering greater flexibility for the end user. This also simplifies model development, deployment and maintenance. Our code is available at https://github.com/ImFusionGmbH/HyperSpace.

Keywords: image segmentation · resolution · hypernetwork · U-Net

1 Introduction

Neural network-based medical image processing has become a standard both in research and clinical settings. This approach relies on statistical learning principles, with the training set's representativeness being crucial for ensuring network robustness in clinical applications. Enhancing this representativeness can be achieved through data augmentation, which generates new training examples to broaden the training set's coverage, or through data standardization, which reduces data variability by applying a consistent pre-processing pipeline during both training and inference phases.

Unlike natural images, medical images typically come with a defined physical voxel spacing (in millimeters), essential for many processing applications. A common pre-processing practice is to use this information to standardize voxel dimensions across different images, ensuring consistency in image analysis. This resampling is crucial for ensuring that convolutional filters consistently interpret the anatomy but is a double-edged sword: reducing resolution from a higher native resolution leads to information loss, while increasing it from a lower native resolution results in the use of computing resources on partly redundant data.

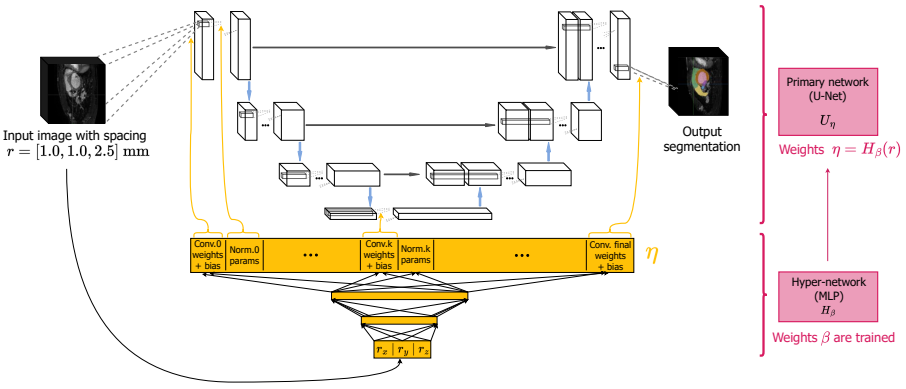

Fig. 1. Illustration of the proposed framework. The hyper-network H_β predicts the primary network's weights η from the image spacing. These weights (and biases) are then dispatched to their corresponding layer in the UNet that performs the segmentation.

We introduce a method to segment images at their native resolution, thus removing the need for resampling, optimizing information usage and minimizing compute requirements. This method is based on a hypernetwork [10], leveraging the segmentation capabilities of U-Nets [20] (Section 3). Hypernetworks are meta-architectures, where one network typically predicts all or parts of the weights of another, based on conditioning variables. In our case, the hypernetwork takes as input the spatial spacing of the image and yields all weights of a segmentation U-Net specific to the chosen resolution. We evaluate the performance of the method on three different datasets and segmentation tasks. Our experiments in Section 4 demonstrate robustness over spatial resolution and that these networks' predictions are comparable to those of models trained for and operated at a fixed image resolution. To provide insights into the internal structure of U-Nets generated from hypernetworks, we compare layers and models using an activation similarity measure. We finally discuss in Section 5 the impact of our work from a methodological and practical perspective.

2 Related work

In the more traditional image processing literature, variable resolution processing was mainly considered from the low-resolution end via the partial volume segmentation issue [25]. In more recent deep processing approaches, to alleviate heterogeneous or unknown image spacings in medical image segmentation, scale equivariance has been incorporated in U-Nets, the reference architecture in medical image segmentation [21,27,29]. Such methods typically add scale as a supplementary dimension and apply $(D+1)$-dimensional convolutions (with D the data dimension) before reducing the scale dimension. Those approaches have however limitations: the scale dimension is discretized and typically considered isotropic as the size of the additional dimension grows exponentially with the

scale dimension. More fundamentally, scale-equivariance is not always desirable, as the structure size can be an essential feature (e.g. vertebrae).

An alternative approach [18] processes the image in Fourier space which allows avoiding the resampling operation but not the scaling discretization, and is almost never used in medical image segmentation. Vision transformers [7] are popular architectures in medical image segmentation [28]. Using world coordinates positional embeddings would make the attention mechanism spacing-adaptive. Yet, the initial patch embedding, the integration within a U-Net architecture [5] or the use of window attention [4] all depend on the image grid.

Another set of methods rely on statistical shape models. Indeed, such models are typically defined with a high node density [23] or even continuously [22]. An agreement between the prior statistical shape model and the observed image can then be found at the native image resolution. Such models tend to be very robust but not extremely accurate, especially in settings of anatomical variety.

Our method is based on hyper-networks [10] which have received comparatively little attention in medical image processing. They were first used in the context of medical image registration [11] as a way to transfer the responsibility to the end-user to tune the regularization regime of the registration. A softer version was presented in [16] where only the normalization layers' parameters are predicted by the hyper-network. Both ideas inspired the present work, but the setting of [11] is more closely related as we predict the full set of resolution-specific parameters of segmentation U-Nets.

To summarize, our key contributions include:

1. The utilization of hypernetworks for spacing-adaptive segmentation.
2. The demonstration of this approach's effectiveness across three distinct datasets.
3. The examination of properties of U-Nets using centered kernel alignment.
4. Enhanced deployment flexibility of U-Net models.

3 Method

3.1 HyperSpace: Spacing-adaptive hypernetworks

Our framework integrates two networks: a standard U-Net architecture segmentation network, U_η, and a hypernetwork, H_β, that predicts the U-nets' weights. In our study, H_β is designed as a straightforward multi-layer perceptron (MLP). This MLP accepts the image voxel resolution in millimeters, $r \in \mathbb{R}^d$ (where d represents the data dimension, e.g. 2 or 3), as its input and produces the weights for the U-Net, $\eta = H_\beta(r)$. Thus, this enables our segmentation framework to adapt to the resolution of the input image. Figure 1 illustrates the overall pipeline.

During training only β is optimized:

$$\beta^* = \arg\min_\beta \mathbb{E}_{(r,\mathcal{X}_r,\mathcal{Y}_r)\sim\mathcal{T}} \left[\mathcal{L}\left(U_{H_\beta(r)}(\mathcal{X}_r), \mathcal{Y}_r)\right)\right] \tag{1}$$

Table 1. Overview of the three public datasets used for our experiments.

Dataset	Segmented Structures	Modality	Training Data	Test Data	Spacing Range (mm)
BRATS	Brain tumour core	T1-CE	900	25	$1.0 \times 1.0 \times 1.0$
SPIDER	Vertebrae, spinal cord	T1	447	10	$[1, 4.8] \times [0.2, 1.2]^2$
MM-WHS	7 Cardiac structures	T2	20	4	$[0.8, 1.1]^2 \times [0.9, 1.6]$

where $\mathcal{T}$ represents the distribution of training data, $\mathcal{X}_r$ and $\mathcal{Y}_r$ are an image and its corresponding label map with voxel resolution r, respectively. $\mathcal{L}$ denotes a supervision loss, in our case a combination of Dice and cross-entropy loss.

In contrast to conventional settings, the expectation is defined over images, corresponding label maps, *and* voxel spacings. These additional dimensions require adequate representation within the training set, further complicating the assembly of a representative dataset. Given the typically limited size of medical image datasets, we confront this challenge by employing a dedicated data augmentation strategy: each training batch is artificially resampled to a random voxel spacing, selected from the dataset's overall spacing range.

3.2 Network analysis using network activation alignment metrics

We aim to study internal properties of U-Net instances generated by hypernetworks, going beyond performance characteristics. In particular, we investigate the similarity of activations across layers and networks using Centered Kernel Alignment (CKA) [13], which is a similarity measure based on the Hilbert-Schmidt Independence Criterion (HSIC) [9] assessing non-parametric independence between random variables. While there are theoretical concerns of CKA [6], it has empirically shown consistency across varying network initializations.

HSIC evaluates similarity by comparing the squared Hilbert-Schmidt norm of the cross-covariance operator in activation spaces. CKA generalizes it by introducing invariance to isotropic scaling. Let $X \in \mathbb{R}^{n \times p_1}$ and $Y \in \mathbb{R}^{n \times p_2}$ represent centered matrices of neural activations for the same n examples but with, in general, different activation counts, p_1 and p_2, respectively. Linear CKA is defined as $\mathrm{CKA}(X, Y) = \frac{\mathrm{HSIC}(K,L)}{\sqrt{\mathrm{HSIC}(K,K)\mathrm{HSIC}(L,L)}}$ with $K = XX^T$, and $L = YY^T$, the activation covariance matrices. For discussion of these metrics, please see [6].

4 Experiments

4.1 Datasets and Baselines

Our experiments utilize public datasets of 3D MRI scans across three distinct datasets and segmentation tasks: BRATS 2021 [1], SPIDER [8], and MM-WHS [30]. We established all training and testing splits randomly. The characteristics of these datasets are summarized in Table 1.

Table 2. Mean Dice score (std) on the three test sets for different resolution subspaces. The first interval considered corresponds to the expected resolution space. The second interval is centered around the datasets' median resolution. The third reported results consider images at their native spacing.

Datasets	**Methods** (see Section 4.1)			
	FS	FSNR	AS	HyperSpace (ours)
BRATS $[0.5, 3.5]^3$	0.92 (0.06)	0.56 (0.36)	0.89 (0.13)	0.91 (0.09)
BRATS $[0.8, 1.2]^3$	0.93 (0.06)	0.88 (0.13)	0.85 (0.15)	0.91 (0.07)
SPIDER $[1, 5] \times [0.2, 1.5]^2$	0.89 (0.01)	0.18 (0.03)	0.87 (0.07)	0.88 (0.07)
SPIDER $[3, 3.5] \times [0.4, 0.8]^2$	0.90 (0.01)	0.20 (0.13)	0.88 (0.03)	0.90 (0.02)
SPIDER native	0.91 (0.01)	0.14 (0.01)	0.86 (0.08)	0.87 (0.08)
MM-WHS $[0.5, 3.5]^3$	0.74 (0.08)	0.20 (0.03)	0.75 (0.01)	0.79 (0.04)
MM-WHS $[0.7, 1.1]^2 \times [1, 1.4]$	0.74 (0.09)	0.72 (0.01)	0.77 (0.07)	0.77 (0.07)
MM-WHS native	0.74 (0.11)	0.73 (0.08)	0.76 (0.08)	0.79 (0.06)

We aim at investigating whether (i) our model demonstrates robustness across large resolution ranges, (ii) outperforms standard data augmentation, and (iii) provides segmentations as accurate as models trained at a fixed resolution but at a lower computational cost. We therefore compared our method **HyperSpace (HS)** to the following baselines:

FixedSpacing (FS) : A U-Net resampling images to a fixed resolution r both at training and inference time (data harmonization). This corresponds to the standard practice. We note that recent approaches only employ data harminization during inference time [2].

FixedSpacingNoResampling (FSNR) : A U-Net resampling images to a fixed resolution r only during training (and processing images at their native resolution at inference). This is a "dummy" baseline demonstrating the importance of considering voxel spacing.

AugmentSpacing (AS) : A single U-Net trained with images at various voxel resolutions which is hence able to process images at native resolution at inference time (data augmentation). This would be the most straightforward solution to deal with images at native resolution.

All segmentation models share the very same architecture corresponding to a 4 levels UNet with 3 convolution block per level, ReLU activations and instance normalization. The hypernetwork is a fully connected network with only 3 hidden layers, ReLU activations and a custom final activation ($: x \longrightarrow tanh(x) * 5$) as a way to constraint the norm of the output weights. All models have been trained for 250000 iterations under identical settings with the same dataset and a combination of Dice and cross-entropy loss. The training procedure is very standard and all additional details are available in our code to be released.

4.2 Performance comparisons

We evaluated the different models on the held-out test sets, across various resolution intervals via Monte Carlo sampling. The findings are presented in Table 2.

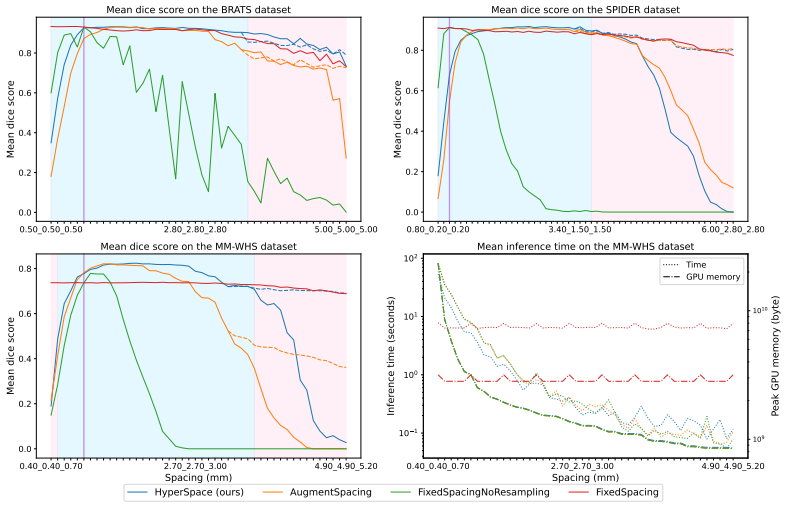

Fig. 2. Mean Dice score for all 3 datasets. On the bottom right, the inference runtime and peak GPU memory usage on the MM-WHS dataset are reported. The blue-shaded region correspond to expected resolution range, the pink-shaded region correspond to resolutions not seen during training. The purple vertical line indicates the resolution at which FS and FSNR were trained.

Dice scores were computed at the simulated native resolution referred to as "working resolution". As an initial sanity check, the poor performances of FSNR on larger resolution intervals highlights the importance of spacing considerations. All 3 methods FS, AS and HyperSpace perform comparably on these 3 tasks. HyperSpace seems to yield better performances on the cardiac dataset where the task is presumably more complex due to the number of target structures. However, the performance of FS comes with a significant and constant computational cost, as shown in Figure 2 (bottom right), for processing time and GPU memory usage. Conversely, the experiment shows that both of these requirements can be decreased by an order of magnitude by processing images at lower resolution. Importantly, as these measures include the full processing pipeline, including the hyper-network forward pass for HyperSpace, the additional processing cost of using a hypernetwork is negligible compared to the segmentation network's inference.

To further investigate the behavior of the different models, performances were densely evaluated along several resolution segments, potentially going beyond the resolutions seen during training (see Figure 2). The hypernetwork delivers more consistent performance throughout the full expected range of resolutions, especially notable in the BRATS and MM-WHS datasets. Indeed, we observe similar performances between AS and HyperSpace at the higher resolution end, while performances tend to diverge in favor of our method on the lower resolution side. We hypothesize that data augmentation is sufficient to cover smaller spacing

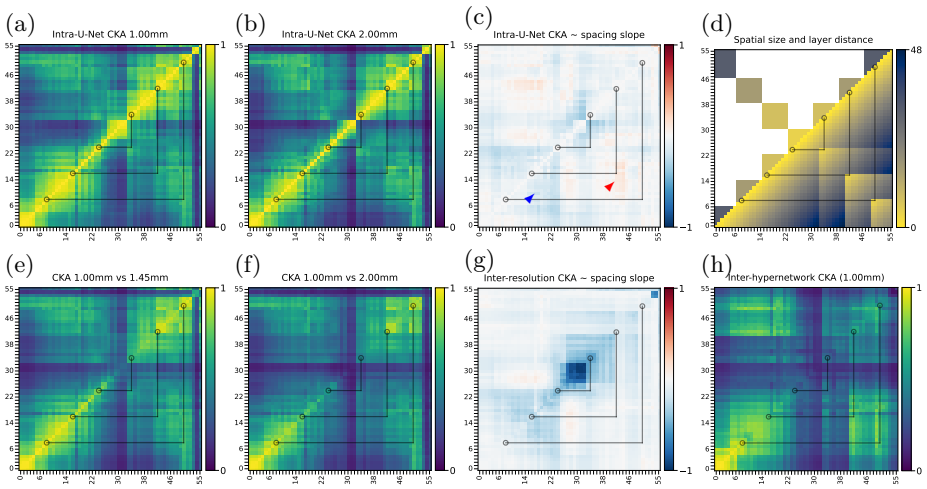

Fig. 3. CKA analysis across convolution and nonlinearity layers, within resolution-specific U-Net networks (a,b) and across networks of 1mm and a coarser resolution network generated from the same hypernetwork (e,f). Plots (c,g) show the rate of change of CKA with spacing for a linear model CKA $\tilde{s}$pacing for the within- (c) and inter resolution-specific U-Net scores (g). For comparison, (h) shows the CKA for two U-Nets with identical spacing but from different hypernetworks. The distance between layers is shown in the lower triangle of (d) with lines indicating skip connections between U-Net branches and the upper triangle shows areas of identical spatial feature dimensions.

ranges but the network weights should be adapted when the voxel size of the target structure varies significantly between images.

However, these graphs also show that the performances of networks processing data at working resolution collapse on the highest resolution end. This is presumably due to the primary network's architecture being too shallow to process such high-resolution images, representing a limitation of the proposed solution as the hypernetwork currently only impacts the weights of the primary network. Being able to predict the weights, as well as modifying the architecture could allow to reliably cover larger resolution intervals. Furthermore, while the hypernetwork manages to extrapolates well to unseen resolutions on the BRATS dataset, the performance rapidly drops on the two others. This decrease is however not as pronounced as for AS. These limitations could be mitigated by resampling, at test time, to the closest training resolution; as the dotted lines in Figure 2 demonstrate, this indeed improves accuracy, matching that of FS. Similarly, at very high resolution, the data can be downsampled to a resolution coherent with the segmentation network architecture.

4.3 Internal representation analysis of generated U-Net networks

Using 437 images of the BRATS dataset for network activation generation, we compute CKA maps (see Section 3.2) for different U-Nets using activations after

convolution and nonlinearity layers. To study the internals of hypernetworks, resolutions of images and networks are varied jointly between 0.94 and 2mm in 10 steps. Via resampling and cropping, we ensure that the extracted activation tensors used for CKA correspond to the same image and feature-space areas irrespective of image resolution. For compute reasons, we chose crops that correspond to a $32 \times 32 \times 32$ image volume at 1mm resolution.

We observe that intra-network CKA scores of U-Nets from the same hypernetwork are relatively stable over resolutions (Figure 3 a,b,c) and hypernetwork seed (not shown). In coarser resolution networks, layers between encoding and decoding branches connected through skip connections are marginally more similar (red arrow in c), while information is less well preserved across layers within the encoder and decoder branches (blue arrow in c). While this resolution-dependent behavior is highly complex, [14] indicates it should converge in expectation and in the infinite width regime, to a neural network Gaussian Process with a covariance derived from the scalar product of spacings.

When comparing activations between 1mm and other resolution-specific U-Net instances from the same hypernetwork, central layers of the U-Net are the least similar across resolutions (Figure 3 g, center) indicating different information aggregation patterns in the deeper layers of the U-Net.

U-Nets from differently seeded hypernetworks share similar information in early layers (Figure 3 h) but compared to intra-network CKA and to U-Networks from the same hypernetwork (Figure 3 a), exhibit little one-to-one layer correspondence in the spatially coarse stages.

To summarize, we can smoothly vary the resolution inputted to the hypernetwork, and the generated models, although resolution-specific, still remain closely related to each other. We hypothesize that this structure granted HyperSpace a similar convergence speed compared to other baselines despite recent evidence that HN are harder and take longer to train [17]. Furthermore, to the best of our knowledge, we demonstrate the first CKA-based analysis of the commonly used U-Net architecture in part thanks to the well-behaved hypernetwork output space.

5 Conclusion

In this work, we investigated the use of hyper-networks in the context of medical image segmentation as a unified model to process images at diverse spatial resolutions. We demonstrated on three datasets that a single hypernetwork can generate competitive U-nets for any input spacing. The additional cost of the proposed method is negligible and offers large computational and GPU memory benefits in low-resolution imaging settings. We find that, the generated U-Nets, while still exhibiting resolution-specific behaviors in the deeper layers, also show a certain consistency across resolutions in other parts.

This work opens several further research avenues. We first suggest validating whether our findings that hypernetworks perform as well as fixed resolution networks, translate to other training strategies such as nnUNet [12]. Then, ex-

tending the capabilities of the hyper-network for instance by using size adaptive convolution kernels [19] or hyper-convolutions [15] in the segmentation UNet could allow to robustly cover larger resolution intervals. More generally, hypernetworks can be used to condition the processing of images on other image properties (e.g. contrast, patient info, etc.)

Beyond purely methodological considerations, we believe this generic and simple yet effective idea can have a practical impact. First, the proposed method can be leveraged when spacing information is missing or incorrect, and needs to be adjusted by the user at inference time (e.g. X-ray systems or poorly calibrated images [24,3]). Most importantly, such hypernetworks could be used as a U-Net factory, allowing the deployment of segmentation models that can be adjusted at inference time to the hardware capability or image resolution. Their flexibility alleviates the need for training distinct models per spacing (see, for instance, the popular TotalSegmentator models [26]) and would democratize access to publicly released models, irrespective of computational resources.

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

References

1. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F.C., Pati, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. arXiv e-prints arXiv:2107.02314 (Jul 2021). https://doi.org/10.48550/arXiv.2107.02314
2. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E.: Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining. Medical Image Analysis **86**, 102789 (2023). https://doi.org/https://doi.org/10.1016/j.media.2023.102789
3. Boese, C.K., Wilhelm, S., Haneder, S., Lechler, P., Eysel, P., Bredow, J.: Influence of calibration on digital templating of hip arthroplasty. International Orthopaedics **43**, 1799–1805 (2019)
4. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds.) Computer Vision – ECCV 2022 Workshops. pp. 205–218. Springer Nature Switzerland, Cham (2023)
5. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation (2021)
6. Davari, M., Horoi, S., Natik, A., Lajoie, G., Wolf, G., Belilovsky, E.: Reliability of cka as a similarity measure in deep learning. arXiv preprint arXiv:2210.16156 (2022)
7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)

8. van der Graaf, J.W., van Hooff, M.L., Buckens, C.F.M., Rutten, M., van Susante, J.L.C., Kroeze, R.J., de Kleuver, M., van Ginneken, B., Lessmann, N.: Lumbar spine segmentation in mr images: a dataset and a public benchmark (2023)
9. Gretton, A., Bousquet, O., Smola, A., Schölkopf, B.: Measuring statistical dependence with hilbert-schmidt norms. In: International conference on algorithmic learning theory. pp. 63–77. Springer (2005)
10. Ha, D., Dai, A.M., Le, Q.V.: Hypernetworks. In: International Conference on Learning Representations (2017)
11. Hoopes, A., Hoffmann, M., Greve, D.N., Fischl, B., Guttag, J., Dalca, A.: Learning the effect of registration hyperparameters with hypermorph. Machine Learning for Biomedical Imaging **1**, 1–30 (2022). https://doi.org/10.59275/j.melba.2022-74f1
12. Isensee, F., Jaeger, P.F., Kohl, S.A.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 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z
13. Kornblith, S., Norouzi, M., Lee, H., Hinton, G.: Similarity of neural network representations revisited. In: International conference on machine learning. pp. 3519–3529. PMLR (2019)
14. Littwin, E., Galanti, T., Wolf, L., Yang, G.: On infinite-width hypernetworks. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS'20, Curran Associates Inc., Red Hook, NY, USA (2020)
15. Ma, T., Dalca, A.V., Sabuncu, M.R.: Hyper-convolution networks for biomedical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1933–1942 (2022)
16. Mok, T.C.W., Chung, A.C.S.: Conditional deformable image registration with convolutional neural network. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. pp. 35–45. Springer International Publishing, Cham (2021)
17. Ortiz, J.J.G., Guttag, J., Dalca, A.V.: Magnitude invariant parametrizations improve hypernetwork learning. In: The Twelfth International Conference on Learning Representations (2024)
18. Rahman, M.A., Yeh, R.A.: Truly scale-equivariant deep nets with fourier layers. In: Thirty-seventh Conference on Neural Information Processing Systems (2023)
19. Romero, D.W., Bruintjes, R.J., Tomczak, J.M., Bekkers, E.J., Hoogendoorn, M., van Gemert, J.: Flexconv: Continuous kernel convolutions with differentiable kernel sizes. In: International Conference on Learning Representations (2022)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015 (2015)
21. Sangalli, M., Blusseau, S., Velasco-Forero, S., Angulo, J.: Scale-equivariant u-net. In: 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022 (2022)
22. Stolt-Ansó, N., McGinnis, J., Pan, J., Hammernik, K., Rueckert, D.: Nisf: Neural implicit segmentation functions. In: MICCAI 2023. pp. 734–744 (2023)
23. Suinesiaputra, A., Albin, P., Alba, X., Alessandrini, M., Allen, J., Bai, W., Cimen, S., Claes, P., Cowan, B., D'hooge, J., Duchateau, N., Ehrhardt, J., Frangi, A., Gooya, A., Grau, V., Lekadir, K., et al.: Statistical shape modeling of the left ventricle: myocardial infarct classification challenge. IEEE Journal of Biomedical and Health Informatics (99) (2017). https://doi.org/10.1109/JBHI.2017.2652449
24. Tafti, A., Byerly, D.: X-ray radiographic patient positioning (2022), updated 2022 Dec 11

25. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: A unifying framework for partial volume segmentation of brain mr images. IEEE Transactions on Medical Imaging **22**(1), 105–119 (2003). https://doi.org/10.1109/TMI.2002.806587
26. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023). https://doi.org/10.1148/ryai.230024
27. Wimmer, T., Golkov, V., Dang, H.N., Zaiss, M., Maier, A., Cremers, D.: Scale-equivariant deep learning for 3D data (2023)
28. Xiao, H., Li, L., Liu, Q., Zhu, X., Zhang, Q.: Transformers in medical image segmentation: A review. Biomedical Signal Processing and Control **84**, 104791 (2023). https://doi.org/https://doi.org/10.1016/j.bspc.2023.104791
29. Yang, Y., Dasmahapatra, S., Mahmoodi, S.: Scale-equivariant UNet for histopathology image segmentation. In: Geometric Deep Learning in Medical Image Analysis (2022)
30. Zhuang, X., Li, L., Payer, C., Štern, D., Urschler, M., Heinrich, M.P., Oster, J., Wang, C., Örjan Smedby, Bian, C., Yang, X., Heng, P.A., Mortazi, A., Bagci, U., et al.: Evaluation of algorithms for multi-modality whole heart segmentation: An open-access grand challenge. Medical Image Analysis **58**, 101537 (2019). https://doi.org/https://doi.org/10.1016/j.media.2019.101537