JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

3D U-Net is all you need for Multi-Organ Medical Segmentation

Naidu Vani[1], Gudla Jyothi Sai[1], Mogili Sai Sindhuja[1], Midisetti LekhaKumari[1], and Dr.Konda Chaitanya[2]

[1] Student, Department of CSE, Dr. YSR ANU College of Engineering & Technology, Acharya Nagarjuna University, Guntur - 52251, Andhra Pradesh, India
[2] Assistant Professor, Department of CSE, Dr. YSR ANU College of Engineering & Technology, Acharya Nagarjuna University, Guntur - 522510, Andhra Pradesh, India

Abstract. Medical picture segmentation is an essential task in the field of medicine that allows different organs to be precisely identified and ana- lyzed for both therapeutic and diagnostic purposes. Although 2D U-Nets and Fully Convolutional Neural Networks (FCNs) have demonstrated potential in medical picture segmentation, they have serious limitations. FCNs frequently function less than optimally because they have difficulty capturing intricate anatomical features and complicated spatial hierar- chies. Comparably, 2D U-Nets, which handle individual volumetric data slices, are unable to fully utilize the 3D spatial context, leading to seg- mentations that are less precise and fragmented. In this study, we show that the 3D U-Net design is an improved multi-organ medical segmenta- tion architecture. The 3D U-Net efficiently captures spatial dependencies in volumetric data by expanding the U-Net architecture into three dimen- sions, resulting in a more precise and coherent segmentation of several organs. We demonstrate that 3D U-Net improves segmentation perfor- mance by maintaining the three-dimensional anatomical context, which gets around the drawbacks of FCNs and 2D U-Nets. We demonstrate the accuracy and robustness of 3D U-Net in multi-organ segmentation tasks through comprehensive tests on benchmark medical imaging datasets. The outcomes show notable gains in consistency and accuracy of seg- mentation, underscoring the importance of 3D U-Net in raising the bar for medical image analysis. Our findings indicate that 3D U-Net outper- forms typical FCNs and 2D U-Nets in multi-organ medical segmentation by providing the required architecture to produce high-quality, depend- able results.

Keywords: 2D U-Net · Fully Convolutional Networks · Multi-domain learning · Segmentation · 3D U-Net· Multi-organ segmentation

1 Introduction

Health research and clinical practice both depend on image segmentation. Themost popular technique for automatic medical image segmentation has been

using fully convolutional neural networks (FCNs), like U-Net [20]. Adapting a neural network architecture to a particular task or dataset and training it from scratch are necessary steps in creating an efficient segmentation model [15,21,23]. As an illustration, [10] developed a single segmentation CNN for brain datasets acquired using different methods and scanners. Despite their power, the highly specialized nature of these models makes them difficult to adapt to new jobs with unknown information.By introducing a self-adapting framework that can generate independent models for various tasks, [9] advanced the discipline. Hu- man experts, on the other hand, can pick up new duties quickly and adapt their expertise to existing ones. Multi-task segmentation with all organs of interest present in the same image has been studied previously [12, 22]. Nevertheless, we tackle a more practical and difficult case: datasets that only annotate one or a few anatomical features when scanning a certain area of the human body. A similar problem was addressed by [17], who trained a single CNN on three tasks; however, their model was not intended to be generalizable to other tasks. As such, the task of creating a technique for image segmentation that is both efficient and successful continues to be quite difficult.

A universal data representation may exist across different visual domains, according to Bilen et al. [2, 18, 19]. The Visual Decathlon Challenge was pre- sented [3], which aims to concurrently model 10 distinct visual domains, such as handwritten characters, internet photos, doodles, planktons, and more [18]. This idea, known as ”multi-domain learning,” was accomplished by pre-training a model on ImageNet and then adding parallel residual adapters to it. But the only thing they work on is image classification. This raises the question: ”Is it possible to build a single neural network that can handle tasks related to medical segmentation across multiple domains?

We take inspiration for this objective from earlier research [3,7], especially [7], which has won the Visual Decathlon Challenge thus far. Depthwise separable convolution was proposed by [7] as a solution to the structural heterogeneity of different domains that was ignored by [19]. Depthwise separable convolution splits the process into two sequential steps: depthwise convolution applies an independent filter to each input channel first, and pointwise convolution linearly combines the outputs across all channels at each spatial location. This is in contrast to standard convolution, which processes spatial and channel-wise com-putations simultaneously.The core of their multi-domain network is composed of a single pointwise convolution shared by all domains after a series of parallel channel-wise convolutions, each devoted to a particular domain. This design in- sight is that shared cross-channel inter-dependencies are identified by pointwise convolutions, whilst domain-specific spatial patterns are efficiently captured by channel-wise convolutions. To avoid confusion with the depth dimension of pic- ture volumes, in this paper, we refer to ”depthwise separable convolution” as simply ”separable convolution” and ”depthwise convolution” as ”channel-wise convolution.”

[3] https://www.robots.ox.ac.uk/~vgg/decathlon/

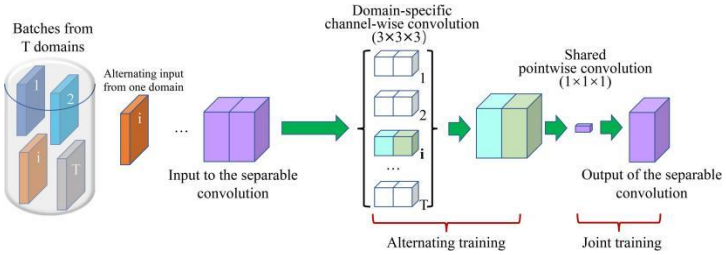

Fig. 1. Domain adapter with separable convolution foundation

Our research suggests a universal architecture for multi-domain medical im- age segmentation, building on the idea of separable convolution. The basic idea is straightforward but effective: we first create a base network with either V- Net [15] or 3D U-Net [4, 20], or a base network using [7] as inspiration. We then replace any 3X3X3.standard convolution with a stride of 1 with separable convo- lution. Our method, however, is very different from [7] in a number of important ways: First of all, their emphasis is on picture categorization, whereas ours is on segmentation of images. Second, we train across domains simultaneously to pro- duce the final model, while they achieve the ultimate multi-domain architecture through a three-step approach comprising pre-training on ImageNet, transfer- ring weights, and domain-specific training.Finally, by adding new channel-wise convolutions, we improve our universal network's adaptability to new domains. This is the first time, as far as we know, that an expandable universal networkhas been created for multi-domain medical image segmentation.

2 Methods

Problem definition:

Consider a set of T image domains, $\{I_1, I_2, \cdots, I_T\}$, where domain I_t is made up of two paired image spaces of X_t, Y_t . The space of the input image is $X_t \in \mathrm{R}^{C_t \times I \times H \times W}$, and the space of the output image, or masks of segmentation, is $Y_t \in \mathrm{R}^{C_t' \times I \times H \times W}$. The spatial dimensions are denoted by the variables I, H, and W. The numbers of imaging modalities and segmentation classes unique to each domain are denoted by C_t and C_t'.Our global network has both shared and domain-specific parameters to function properly across all domains. Let ϑ_m represent domain-specific parameters for domain I_t, and let ϑ_n represent parameters shared by all domains universally. The output $\hat{Y}$ of the neural network $F(X)$ is

$$\hat{y}_{t,i} = F(x_{t,i}; \vartheta_m, \vartheta_n).$$

, assuming $\{x_{i,t}, y_{i,t}\}$ as the i^{th} training pair of domain I_t.

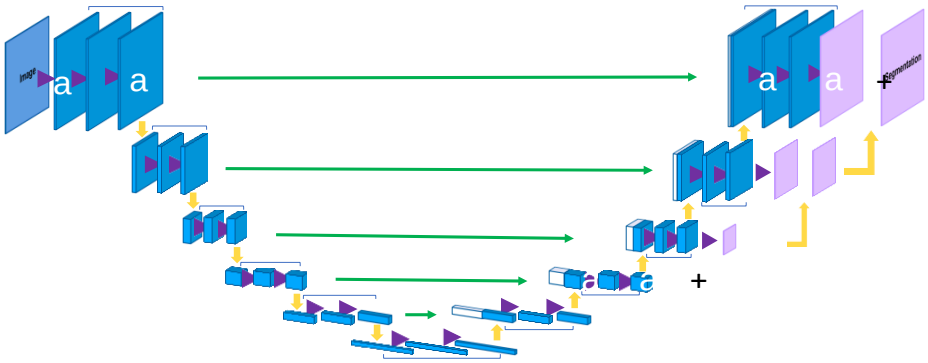
16
32
64
128
256
512
Residual Connection
1×1×1 convolution
Adapter
Down Sampling
Up Sampling
Concatenation
Multi-level segmentations

Fig. 2. The proposed 3D U-Net Architecture(3D U-Net).

Domain Converter:

The domain converter [8], which is essential to the functioning of our universal network, is based on separable convolution rather than conventional convolution and includes both shared and domain-specific parameters.
The output tensor $\hat{O}$ $\mathrm{R}^{X' \times I \times H \times P}$ $\in$ is obtained by applying X' filters w $\mathrm{R}^{3\times 3\times 3\times \times X'}$ $\in$ on the input in parallel and concatenating the X' output featuremaps.This is the usual convolution applied to an input tensor U $\mathrm{R}^{X\times I\times H\times P}$,with filter W $\mathrm{R}^{3\times 3\times 3\times X\times X}$. A simple computation yields the total number offilter parameters in the aforementioned filters: 27 * X * X'.Furthermore, whiletraining the models independently for the T domains, the number of parameters increases T times!
The computation in separable convolution can be divided into two stages that follow one another. The first stage concatenates the X output feature maps by applying X channel-wise filters p $\mathrm{R}^{3\times 3\times 3}$ to every input channel concurrently. Each domain in this case has its own set of channel-wise filters. The final feature maps of X' channels are then output in the second phase by applying x' point-wise filters p $\mathrm{R}^{1\times 1\times 1\times C}$. The point-wise filters are shared by all domains in this case. A short computation indicates that 27 * X * T + X * X'.is thetotal number of weights in the aforementioned filters. Fig. 1 shows how to put together the common point-wise convolution with the domain-specific channel- wise convolutions to create a domain adapter.

Model Architecture(3D U-Net):

Our universal network architecture is based on a fundamental network com- prising six elements, as illustrated in Fig. 1: (1) input; (2) encoder path; (3) bottleneck block; (4) decoder path; (5) deep supervision branch; and (6) output. The number of imaging modalities and classifications of distinct domains may impact the input and output channels. The input layer typically utilizes 16 filters.

There are five layers across the encoder and decoder paths, albeit at different resolutions. Within each level, a residual connection is applied. In the decoder path, skip connections retain more contextual information from the encoder counterpart (Ronneberger et al., 2015). Motivated by [11], we improve the final localization performance by introducing a deep supervision branch alongside thedecoder path via the element-wise sum of multi-level segmentation maps.Any standard 2 2 3 convolution with a stride of 1 is replaced with the domain above adapters and integrated into the basic network to construct the universal network. The proper details of the implementation are given by the table1

Loss function:

Since the unweighted sum of cross-entropy loss and Dice loss [16] has been shown to be reliable for a variety of medical image segmentation tasks [13], we utilized it as the final loss function. Let *SR, GT* stand for the segmentation result and ground truth, respectively, in this instance. For each voxel *i*, the

symbols sr_i, gt_i represent the expected segmentation and ground truth, respectively. The quan- tity of voxels in image *I* is denoted by *N* . One may define binary cross-entropy loss using

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[sr_i \log gt_i + (1 - sr_i)\log(1 - gt_i)]$$

and dice loss is defined by

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} sr_i y_i}{\sum_{i=1}^{N}(gt_i)^2 + \sum_{i=1}^{N}(sr_i)^2}.$$

The final loss *L* is defined by

$L = L_{BCE} + L_{Dice}$.

Metrics of Evaluation:

We employed the Dice Similarity Coefficient and Normalized Surface Distance (NSD) [6] to quantitatively assess the segmentation outcomes, according to the suggestions in Metrics Reloaded [14]. DSC is a region-based segmentation metric
[14] that measures how well expert annotation masks and segmentation results match in the regions that are described by

$$DiceSimilarityScore(GT, SR) = \frac{2|GT \cap SR|}{|GT| + |SR|}$$

A boundary-based metric called NSD [6] seeks to assess the boundary consensus between segmentation results and expert annotation masks at a certain toler- ance, which is determined by

Table 1. Details of Model Architecture.

Layer (type)	Output Shape	#Parmaters
InputLayer	(64, 64, 64, 3)	0
Conv3D	(64, 64, 64, 16)	1312
Dropout	(64, 64, 64, 16)	0
Conv3D	(64, 64, 64, 16)	6928
MaxPooling3D	(32, 32, 32, 16)	0
Conv3D	(32, 32, 32, 32)	13856
Dropout	(32, 32, 32, 32)	0
Conv3D	(32, 32, 32, 32)	27680
MaxPooling3D	(16, 16, 16, 32)	0
Conv3D	(16, 16, 16, 64)	55360
Dropout	(16, 16, 16, 64)	0
Conv3D	(16, 16, 16, 64)	110656
MaxPooling3D	(8, 8, 8, 64)	0
Conv3D	(8, 8, 8, 128)	221312
Dropout	(8, 8, 8, 128)	0
Conv3D	(8, 8, 8, 128)	442496
MaxPooling3D	(4, 4, 4, 128)	0
Conv3D	(4, 4, 4, 256)	884992
Dropout	(4, 4, 4, 256)	0
Conv3D	(4, 4, 4, 256)	1769728
Conv3DTranspose	(8, 8, 8, 128)	262272
Concatenate	(8, 8, 8, 256)	0
Conv3D	(8, 8, 8, 128)	884864
Dropout	(8, 8, 8, 128)	0
Conv3D	(8, 8, 8, 128)	442496
Conv3DTranspose	(16, 16, 16, 64)	65600
Concatenate	(16, 16, 16, 128)	0
Conv3D	(16, 16, 16, 64)	221248
Dropout	(16, 16, 16, 64)	0
Conv3D	(16, 16, 16, 64)	110656
Conv3DTranspose	(32, 32, 32, 32)	16416
Concatenate	(32, 32, 32, 64)	0
Conv3D	(32, 32, 32, 32)	55328
Dropout	(32, 32, 32, 32)	0

Conv3D	(32, 32, 32, 32)	27680
Conv3DTranspose	(64, 64, 64, 16)	4112
Concatenate	(64, 64, 64, 32)	0
Conv3D	(64, 64, 64, 16)	13840
Dropout	(64, 64, 64, 16)	0
Conv3D	(64, 64, 64, 16)	6928
Conv3D	(64, 64, 64, 3)	51
Total params:	5645811	(21.54 MB)
Trainable params:	5645811	(21.54 MB)

where $B^{(\tau)}_{\partial GT} = \{a \in R^3 \mid \exists \tilde{a} \in \partial GT, ||a - \tilde{a}|| \leq \tau\}$, $B^{(\tau)}_{\partial SR} = \{a \in R^3 \mid \exists \tilde{a} \in \partial SR, ||a - \tilde{a}|| \leq \tau\}$ represent the segmentation surface at tolerance τ and the expert annotation mask's boundary region, respectively. We set the tolerance τ to 2 in this work.

3 Experiments and Results

In this section, we do comprehensive experiments to evaluate the 3D U-Net that has been suggested for multiple organ segmentation in medicine. We look into three different approaches: We offer three models: (1) Liberal models, which train the basic network for each base domain separately to emulate conventional approaches; (2) Unified models, which train a single basic network on all base domains concurrently to investigate whether all parameters can be shared across all domains; and (3) Global model, which is our ultimate objective and trains a universal architecture with all base domains concurrently. The first two tech- niques are regarded as extreme examples and established the benchmark for varied organ segmentation. Additionally, we evaluate the shared model's and the universal model's generalizability on a new domain.

Dataset Curation:

We make use of six publicly available datasets from the [1]-introduced Medical Segmentation Decathlon challenge [5]. To train the universal model, the first five datasets are used as base domains. The sixth dataset, which is used to assess the universal model's adaptability, is regarded as a new domain. Table 2 displays the fundamental features of these datasets. 80% of the samples in each dataset are randomly picked for training, while the remaining 20% are used for testing.

4

Preprocessing:

In terms of voxel spacing, image size, and modality, the datasets are extremely varied. Preprocessing steps are carried out as follows: (1) To reduce the compu- tational load, all images of patients are trimmed to the desired shape of (depth, width, height) with (64,64,64) values;(2) for each patient, the image is clipped to the [1,100.0] of the pixel percentiles of the intensity values of the entire image, followed by standard normalization with the mean and standard deviation of the image for each

modality; and (4) The image is augmented with many data augmentation techniques like random clipping, random flipping and random ro- tations with the help of batch generators.
We employ randomly selected patches from the entire image to overcome the GPU memory constraints during training. On the other hand, we use a sliding window technique with a half-patch stride over the image during inference. Everyinput batch for both the universal and shared models is made up of two 2^7

$\times 2^7 \times$

[4] https://decathlon.grand-challenge.org/

Table 2. Basic characteristics of the datasets. [8]

Task	Modality	Training Size	Shape of Image	Voxel spacing
Task01 Heart	MRI	20	0×320	1.37×1.25×1.25
Task02 Liver	CT	131	2×512	1)×(0.357≈1)
Task03 Hippocampus	MRI	260	(259)×(31≈43)	1×1×1
Task04 Prostate	MRI	32	6≈384)×(256	)×(0.6≈0.75)
Task07 Pancreas	CT	281	2×512	0.977)×(0.605≈0.977)
Task09 Spleen	CT	41	2×512	5≈0.977)×(0.535

2^7-sized patches. These batches undergo six different downsampling procedures. We adjust the input patch size and resolution levels for the independent models across many domains while accounting for image size to optimize the utilization of computational resources. To have the patch size with the same aspect ratio as the median shape, we alternate between the input patch size and batch size if the median shape is less than 2^7 2^7 2^7.To prepare the patches for the shared model and universal model, we specifically extract a patch that is the same size as the independent model and then resize it to the objective patch size indicated above.

Implementation details:

The network is implemented on Google Colab[5], in TensorFlow 2.15.0 using an NVIDIA T4 Tensorcore GPU. Deep learning model training is made easier with Google Colab, a free cloud-based platform that runs Python code and gives users access to strong computational resources like GPUs and TPUs.
Several essential elements and parameters are included in the network train-ing process:

– **Weight Decay**: 10^{-5} is the weight decay used. By punishing large weights, the regularization approach known as weight decay helps minimize over- fitting. The network is encouraged to retain lower weights by adding this penalty term to the loss function.
– **Learning Rate**: For the training phase, a preset learning rate of 10^{-4} is used. One important hyperparameter that sets the step size for each iteration as it approaches the loss function minimum is the learning rate. The model parameters are updated gradually when the learning rate is 10^{-4}, which aids in establishing stable convergence.
– **Optimizer**: The ADAM (Adaptive Moment Estimation) optimizer is uti- lized. Stochastic gradient descent is extended by ADAM, which calculates adaptive learning rates for every parameter. It is particularly well-suited for managing sparse gradients on noisy issues because it combines the benefits of AdaGrad and RMSProp, two additional extensions of stochastic gradient descent.

– **Batches and Epochs**: A single full iteration across the whole dataset is called an epoch. An epoch for this network is made up of eight batches. A batch is a portion of the dataset that is used in a single forward and

[5] https://colab.research.google.com/

backward pass to train the network. Training in batches facilitates faster convergence by allowing the model to change its parameters more often than when utilizing the complete dataset in each epoch.

– **Testing Data**: The testing data served as the foundation for the outcomes that were shown. The testing data is used to assess the model's performance after it has been trained on the training set. This assessment aids in com- prehending the model's ability to generalize to fresh, untested data.

Base domain quantitative results:

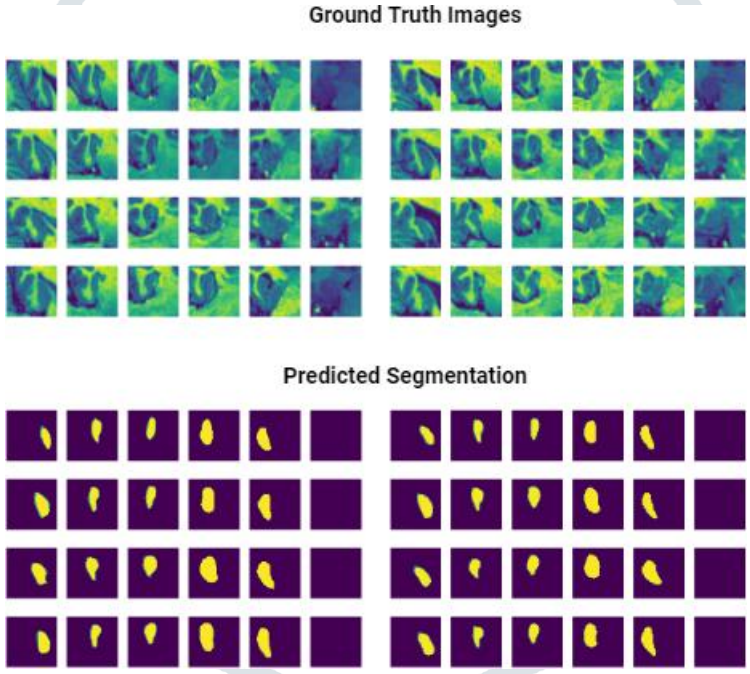

Table 3 lists the three models' average Dice scores for each base domain. The independent models generate the highest mean score overall and score highest across most domains when we compare along the columns. Remarkably, however, both the universal model and the shared model outperform the independent models considerably in terms of Task04 Prostate's transition zone (TZ) and peripheral zone (PZ), and they perform moderately for the bulk of domains. Furthermore, we see that the universal model outperforms the shared model in the pancreas segmentation for Task05 Pancreas. Every domain could be the reason for the performance gain overall.

Fig. 3. Results of the qualitative assessment on the testing set:The ground truth MRI scan from different orientations and its corresponding predicted segmenta-tion

Table 3. quantitative outcomes in relation to the base domains.

	Task01 Heart	Task02 Liver	Task03 Hippocampus		Task04 Prostate		Task05 Pancreas	
(Dice %)	Left atrium	Liver	Anterior	Posterior	PZ	TZ	Pancreas	Mean
Liberal	90.24	92.22	87.72	86.94	56.79	87.68	79.68	86.48
Unified	91.63	90.89	90.27	89.38	66.39	83.38	54.47	80.76
Global	90.48	92.34	88.64	86.54	67.52	86.24	62.08	85.10

Fig. 4. Training and Validation Results

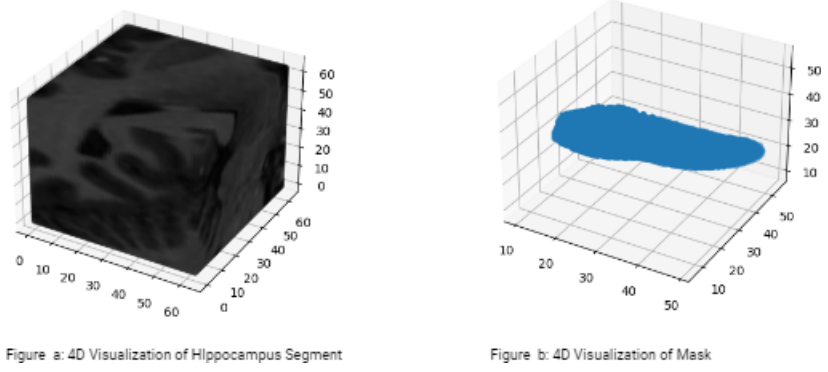

Fig. 5. 4D hippocampus segmentation with its mask

4 Conclusions

In summary, we introduce a novel universal neural network named 3D U-Net for multi-organ segmentation problems, thereby filling the gap in extendable multi- domain learning in picture

segmentation. The experiment results demonstrate that the proposed method can obtain segmentation performance comparable to independent models trained with the standard method by modifying a few parameters. In the near future, a completely developed universal model for multi-organ segmentation can be accomplished, given the constant number of human organs and the regular availability of CT and MRI data. Moreover, the proposed methodology could have broad applicability across other domains, expediting theimplementation of neural networks in therapeutic contexts.

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Land- man, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., van Gin- neken, B., Bilello, M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M.J., Heck- ers, S.H., Huisman, H., Jarnagin, W.R., McHugo, M.K., Napel, S., Pernicka, J.S.G., Rhode, K., Tobon-Gomez, C., Vorontsov, E., Meakin, J.A., Ourselin, S., Wiesenfarth, M., Arbeláez, P., Bae, B., Chen, S., Daza, L., Feng, J., He, B., Isensee, F., Ji, Y., Jia, F., Kim, I., Maier-Hein, K., Merhof, D., Pai, A., Park, B., Perslev, M., Rezaiifar, R., Rippel, O., Sarasua, I., Shen, W., Son, J., Wachinger, C., Wang, L., Wang, Y., Xia, Y., Xu, D., Xu, Z., Zheng, Y., Simpson, A.L., Maier-Hein, L., Cardoso, M.J.: The medical segmentation decathlon. Na- ture Communications **13**(1) (2022). https://doi.org/10.1038/s41467-022-30695-9, http://dx.doi.org/10.1038/s41467-022-30695-9
2. Bilen, H., Vedaldi, A.: Universal representations: The missing link between faces, text, planktons, and cat breeds. arXiv:1701.07275 (2017)
3. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proc. CVPR. pp. 1251–1258 (2017)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Proc. MICCAI. pp. 424–432. Springer (2016)
5. Decathlon Organizers: Medical segmentation decathlon (2018), https:// decathlon.grand-challenge.org/
6. DeepMind: surface-distance. https://github.com/google-deepmind/ surface-distance (2018)
7. Guo, Y., Li, Y., Feris, R., Wang, L., Rosing, T.: Depthwise convolution is all you need for learning multiple visual domains. arXiv:1902.00927 (2019)
8. Huang, C., Han, H., Yao, Q., Zhu, S., Zhou, S.K.: 3d u^2-net: A 3d universal u-netfor multi-domain medical image segmentation (2019)
9. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnU-Net: Self- adapting framework for u-net-based medical image segmentation. arXiv:1809.10486 (2018)
10. Karani, N., Chaitanya, K., Baumgartner, C., Konukoglu, E.: A lifelong learning approach to brain MR segmentation across scanners and protocols. In: Proc. MIC- CAI. pp. 476–484. Springer (2018)
11. Kayalibay, B., Jensen, G., van der Smagt, P.: CNN-based segmentation of medical imaging data. arXiv:1701.03056 (2017)
12. Lay, N., Birkbeck, N., Zhang, J., Zhou, S.K.: Rapid multi-organ segmentation using context integration and discriminative models. In: Proc. MICCAI. pp. 450–462. Springer (2013)
13. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Medical Image Analysis **71**, 102035 (2021). https://doi.org/10.1016/j.media.2021.102035
14. Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M.D., Buettner, F., Christodoulou, E., Glocker,

B., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M.A., Wiesenfarth, M., Kavur, A.E., Sudre, C.H., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Rädsch, T., Acion, L., Antonelli, M., Arbel, T., Bakas, S., Benis, A., Blaschko, M.B., Cardoso, M.J., Cheplygina, V., Cimini, B.A., Collins, G.S., Farahani, K., Ferrer, L., Galdran, A., van Ginneken, B., Haase, R., Hashimoto, D.A., Hoffman, M.M., Huisman, M., Jannin, P., Kahn, C.E., Kain- mueller, D., Kainz, B., Karargyris, A., Karthikesalingam, A., Kofler, F., Kopp- Schneider, A., Kreshuk, A., Kurc, T., Landman, B.A., Litjens, G., Madani, A., Maier-Hein, K., Martel, A.L., Mattson, P., Meijering, E., Menze, B., Moons, K.G.M., Müller, H., Nichyporuk, B., Nickel, F., Petersen, J., Rajpoot, N., Rieke, N., Saez-Rodriguez, J., Sánchez, C.I., Shetty, S., van Smeden, M., Summers, R.M., Taha, A.A., Tiulpin, A., Tsaftaris, S.A., Van Calster, B., Varoquaux, G., Jäger, P.F.: Metrics reloaded: recommendations for image analysis validation. Nature Methods **21**(2), 195–212 (Feb 2024). https://doi.org/10.1038/s41592-023-02151-z, http://dx.doi.org/10.1038/s41592-023-02151-z

15. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proc. 3DV. pp. 565–571 (2016)
16. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation (2016)
17. Moeskops, P., Wolterink, J.M., van der Velden, B.H., Gilhuijs, K.G., Leiner, T., Viergever, M.A., Išgum, I.: Deep learning for multi-task medical image segmenta- tion in multiple modalities. In: Proc. MICCAI. pp. 478–486. Springer (2016)
18. Rebuffi, S.A., Bilen, H., Vedaldi, A.: Learning multiple visual domains with residual adapters. In: Proc. NIPS. pp. 506–516 (2017)
19. Rebuffi, S.A., Bilen, H., Vedaldi, A.: Efficient parametrization of multi-domain deep neural networks. In: Proc. CVPR. pp. 8119–8127 (2018)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomed- ical image segmentation. In: Proc. MICCAI. pp. 234–241. Springer (2015)
21. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: DeepOrgan: Multi-level deep convolutional networks for automated pancreas seg- mentation. In: Proc. MICCAI. pp. 556–564. Springer (2015)
22. Roth, H.R., Oda, H., Hayashi, Y., Oda, M., Shimizu, N., Fujiwara, M., Misawa, K., Mori, K.: Hierarchical 3D fully convolutional networks for multi-organ segmen- tation. arXiv:1704.06382 (2017)
23. Savioli, N., Montana, G., Lamata, P.: V-FCNN: Volumetric fully convolution neural network for automatic atrial segmentation. arXiv:1808.01944 (2018)