TriSegNet: A Lightweight Trilateral Segmentation Network for Aabdominal Organ segmentation

Wei Chen
School and Hospital of Stomatology, Cheeloo College of Medicine, Shandong University
Shandong Key Laboratory of Oral Tissue Regeneration
Shandong Engineering Laboratory for Dental Materials and Oral Tissue Regeneration
No.44-1 Wenhua Road West, 250012, Jinan, Shandong, China
ichenwei@sdu.edu.cn

Xianru Zhang
School of Control Science and Engineering, Shandong University
zhangxianru@mail.sdu.edu.cn

Xu Qiao
School of Control Science and Engineering, Shandong University
qiaoxu@sdu.edu.cn

Abstract

Accurate abdominal organ segmentation from computerized tomography (CT) scans has enormous potential value for clinical diagnosis, treatment planning, and prognosis evaluation. With the successful development of deep learning techniques, many algorithms have been proposed for abdominal organ segmentation and achieved promising results. However, abdominal organ segmentation remains challenging because of the large model size and cost extensive computational resources. In this paper,We propose a novel convolutional neural network (CNN) architecture called Trilateral Segmentation Network (TriSegNet) to solve these problems. The proposed TriSegNet consists of three Paths, namely a Spatial Path to capture rich spatial information, a Context Path to obtain sufficient receptive field and encode rich context information, and a Localization Path to recover the original input size and restore tumor details. Our model is trained on entire input images and can generate the segmentation result for a whole volumetric image in a single pass. Crucially, to make the model less complicated and reduce the number of parameters, we adopt the idea of spatiotemporal-separable 3D convolution and 3D depthwise separable convolution. Extensive experiments are conducted on imaging data provided by the Fast and Low GPU memory Abdominal oRgan sEgmentation (FLARE). Compared with other state-of-the-art CNN-based methods, TriSegNet achieves competitive performance with a small size of 9MB. In the evaluation step, the proposed method achieves a mean liver, kidney, spleen and pancreas dice scores of 0.929, 0.752, 0.828 and 0.652, respectively.

1. Introduction

In the past several decades, many researchers in medical imaging and computer science have done a lot of work for abdominal organ segmentation to assist clinicians in their diagnosis. Most segmentation methods fall into two broad categories: generative probabilistic methods and discriminative methods. Generative probabilistic methods utilize domain-specific prior knowledge about the appearance and spatial distribution of different organs. Specifically, they model the intensities as a posteriori distribution using probability theory, where prior distribution encodes the spatial knowledge of different organs. Compared with the generative methods, the discriminative methods extract a large number of features from raw images and directly model the relationship between these features and the ground-truth segmentation.

In recent years, convolutional neural networks (CNNs) are widely used in computer vision. Neural networks can automatically mine feature representations from raw data in a hierarchical manner, resulting in a more robust and powerful representation than handcrafted features. They have achieved great promising results in many tasks, including image classification [1], objects detection [2] and semantic

segmentation [3]. Following the general computer vision field's success, many works on organ segmentation have achieved significant improvements by adopting the state-of-the-art CNNs.

The above methods are based on 2D convolutions and only learn features by independently processing each 2D slice. However, volumetric medical data, such as abdominal CT, have inherent 3D characteristics. The volumetric images can show richer information than independent 2D slices. Doctors can easily identify some valuable anatomical locations and pathological characteristics from the 3D context. However, it is a challenging task for 2D slices [4]. To fully explore 3D information, Kamnitsas et al.[5] proposed a 3D CNN-based method called DeepMedic for various medical image segmentation tasks. To address the high memory demand and substantial computational burden of 3D convolutions, the authors proposed dividing the input image into patches and utilizing small kernels. They also employed parallel convolutional pathways to extract multi-scale features, with one pathway receiving patches of the original image and another receiving larger patches downsampled to a lower resolution. Furthermore, they employed a conditional random field (CRF) to refine the CNN output.

A drawback of above methods is that the learning process is limited to visible features in each patch, and the network typically cannot learn global features. The fully convolutional network (FCN) proposed by Long et al. [3] can be trained end-to-end for pixel-wise inference. In more detail, the fully connected layers are replaced by convolutions, so CNN can take images of any size as input. Upsampling is introduced to make the output have the same size as the input image. Later, Ronneberger et al. [6] extended FCN architecture for medical image segmentation and proposed an end-to-end encoder-decoder network called U-Net. This network includes a contracting path that captures context information and an expanding path that allows precise localization, with skip connections between them.

Different organs have different shapes and sizes, and the boundaries are often ambiguous. Therefore, it is very important to preserve the resolution of the input image to encode enough spatial information. In addition, a large enough receptive field is also indispensable for encoding high-level semantic context information. So, the critical problem in organ segmentation is how to learn highly discriminative semantic features while preserving high spatial resolution. However, most current methods are struggling to meet these requirements simultaneously.

To address above challenges, we propose Trilateral Segmentation Network (TriSegNet) in the present work. TriSegNet has three paths: a Spatial Path (SP) to capture rich spatial information, a Context Path (CP) to obtain sufficient receptive field and encode rich context information, and a Localization Path (LP) to recover the original input size and restore tumor details. Our network is inspired by Bilateral Segmentation Network (BiSeNet) [7] which is designed for nature image segmentation. We modify and extend BiSeNet to make it suitable for volumetric medical images. Extensive experiments are conducted on the dataset provided by the Fast and Low GPU memory Abdominal oRgan sEgmentation (FLARE). The segmentation results prove the effectiveness of our method with a significantly smaller size.

2. Method

Figure 1 shows the pipeline of TriSegNet, which consists of a Spatial Path, a Context Path and a Localization Path. In the following subsections, we will illustrate the three paths in detail.

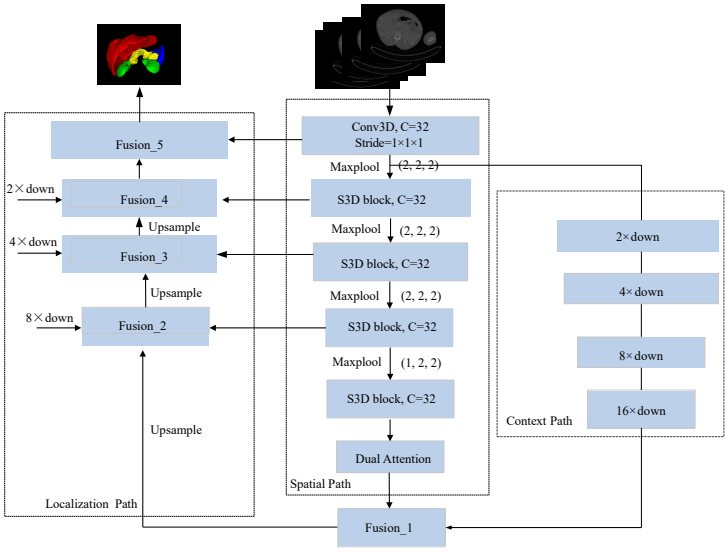

Figure 1. Network architecture

2.1. Spatial Path

As the name states, the Spatial Path is designed to capture rich spatial information. The similar path in BiSeNet [7] consists of only three layers, and each layer contains a convolution with stride of 2. Although rich spatial information can be preserved, the output feature maps may contain too much noise to provide enough high-resolution semantic information [8]. To address this problem, we consider adding more convolutional layers. However, more convolutional layers in the context of 3D convolutions will result in unacceptable high computational cost and memory consumption. To resolve the dilemma, we adopt the idea of spatiotemporal-separable 3D convolutions[9], that is,the information extracted by each 3D convolution can be approximated by two consecutive convolution layers. The fisrt 2D convolution is used to extract spatial features and the second

1D is used to extract temporal features. This separable 3D convolutions are more computationally efficient than standard 3D convolutions.

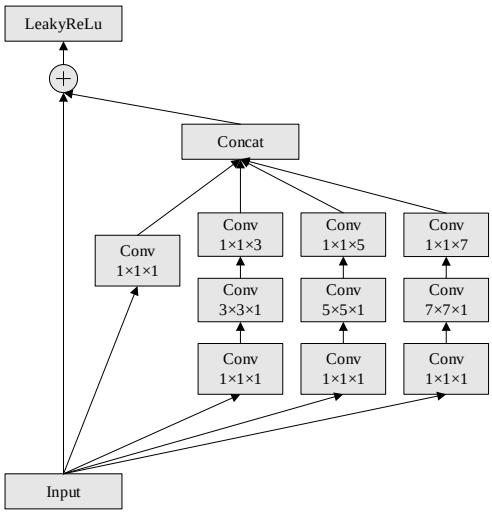

Figure 2. S3D block that utilizes spatiotemporal-separable 3D convolutions in an Inception-ResNet-like architecture.

Since different tumor regions vary significantly in size, it is difficult to choose the optimal kernel sizes for convolutions. The Inception architecture proposed by Szegedy et al. [10] can deal with this problem by performing convolutions of different kernel sizes in parallel and concatenating the resulting feature maps before entering the next layer. Inspired by the advantages of ResNet [11] in training very deep architectures, they later proposed Inception-ResNet [12] that combined the Inception architecture with residual connections and achieved state-of-the-art performance on several challenging tasks. Inspired by the successes of Inception-ResNet, we develop a new Separable-3D (S3D) block that utilizes spatiotemporal-separable 3D convolutions in an Inception-ResNet-like architecture, as shown in Figure 2. Thanks to the use of lightweight spatiotemporal-separable 3D convolutions, we can adopt larger kernel sizes to better encode spatial information. At the end of the Spatial Path, we added a dual attention module [13] to capture rich contextual dependencies based on the self-attention mechanism.

2.2. Context Path

In the task of semantic segmentation, it is widely accepted that a larger receptive field leads to better performance. Some 2D networks use pyramid pooling module [14], atrous spatial pyramid pooling [15] or "large kernel" [16] to capture context features from the entire image. Due to the challenges of organ segmentation, a sufficiently large 3D receptive field is essential for obtaining accurate segmentation. However, the typical operations adopted in 2D CNNs require a large amount of memory and computation in the context of 3D CNNs, which limits their ability to build complex models and effective representations.

In order to obtain a large enough receptive field while ensuring the efficiency of the calculation, we utilize the idea of depthwise separable convolutions. Depthwise separable convolutions have been adopted in several networks and have achieved excellent performance [17],[18], [19], [20]. The core idea of depthwise separable convolutions is to factorize the standard 2D convolution into a depthwise convolution, and a 1×1 convolution called a pointwise convolution. Similar to the networks proposed by Hu et al. [20] and Ye et al. [21], we extend 2D depthwise separable convolutions to 3D. Specifically, we decompose the standard 3D convolution into a 3D depthwise convolution, and a $1 \times 1 \times 1$ convolution called a 3D pointwise convolution, as shown in Figure 1.

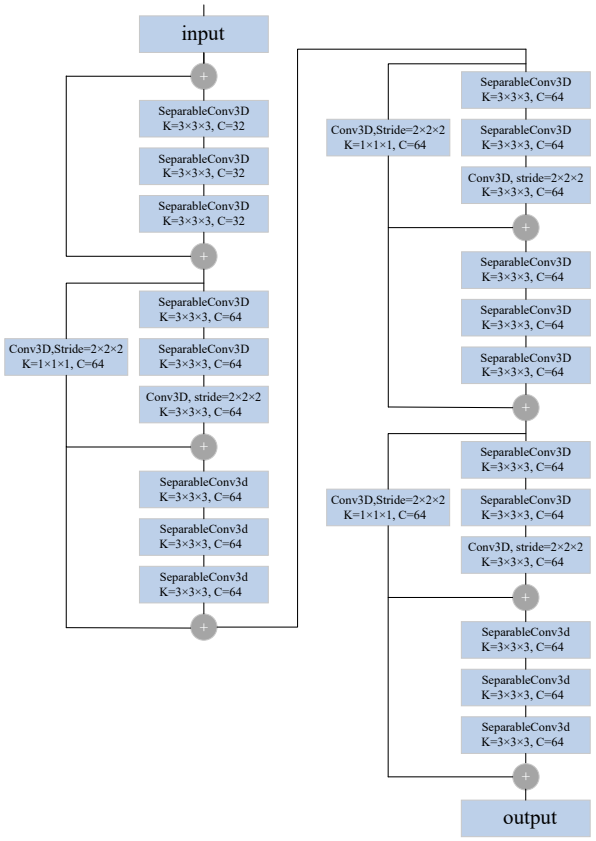

Figure 3. The Context Path. Unless otherwise stated, all layers use kernels of size $3 \times 3 \times 3$ with a stride of 1.

A complete description of the Context Path is given in Figure 3. Similar to Xception Network [19], our architecture is a linear stack of depthwise separable 3D convolutional layers with residual connections. All convolutional layers are followed by instance normalization [22]

and LeakyReLu [23].

2.3. Localization Path

The size of the output feature maps is reduced to 1/8 of the original MRI through the processing of Spatial Path and Context Path, The strategy for recovering the original input size in BiSeNet is to upsample the final feature maps 8 times. We argue that this strategy cannot effectively restores tumor details in organ segmentation task. Most state-of-the-art semantic segmentation networks employ stacked upsampling layers, which are commonly called expanding path or decoder path, to recover the feature maps to the original input size. The spatial information lost due to subsampling can be restored by skipping features of the same resolution from the encoder path to the corresponding decoder path. Inspired by the U-Net architecture, we propose the Localization Path (LP) consisting of four fusion modules to recover the input dimensions. An overview of the Localization Path is shown in Figure 1.

2.4. Preprocessing

We adopted the preprocessing step as the same as 3D nnU-Net [25].The baseline method includes the following preprocessing steps:

- Cropping strategy: None.
- Resampling method for anisotropic data:
  In-plane with third-order spline interpolation, out-of-plane with nearest neighbor interpolation.
- Intensity normalization method:
  First, the dataset is clipped to the [0.5, 99.5] percentiles of the intensity values of the training dataset. Then a z-score normalization is applied based on the mean and standard deviation of the intensity values.

2.5. Proposed Method

- Network architecture details: the architecture of our proposed method has discussed in section 2. detail of each layer, hyper-parameters, such as strides, weights size, etc.
- Loss function: we use the summation between Dice loss and cross entropy loss because it has been proved to be robust [26] in medical image segmentation tasks.
- Number of model parameters: 1158542.

2.6. Post-processing

A connected component analysis of all ground truth labels is applied on training data [25].

3. Dataset and Evaluation Metrics

3.1. Dataset

- A short description of the dataset used:
  The dataset used of FLARE2021 is adapted from MSD [27] (Liver [28], Spleen, Pancreas), NIH Pancreas [29, 30, 31], KiTS [32, 33], and Nanjing University under the license permission. For more detail information of the dataset, please refer to the challenge website and [34].
- Details of training / validation / testing splits:
  The total number of cases is 511. An approximate 70%/10%/20% train/validation/testing split is employed resulting in 361 training cases, 50 validation cases, and 100 testing cases. The detail information is presented in Table 1.

3.2. Evaluation Metrics

- Dice Similarity Coefficient (DSC)
- Normalized Surface Distance (NSD)
- Running time
- Maximum used GPU memory (when the inference is stable)

4. Implementation Details

4.1. Environments and requirements

The environments and requirements of the baseline method is shown in Table 2.

4.2. Training protocols

The training protocols of the baseline method is shown in Table 3.

4.3. Testing protocols

- Pre-processing steps of the network inputs:
  The same strategy is applied as trainging steps.
- Post-processing steps of the network outputs:
  No post-processing step is used.
- If using patch-based strategy, describing the patch aggregation method:
  The same patch-based strategy is applied as nnU-Net [25]. Voxels close to the center are weighted higher than those close to the border, when aggregating predictions across patches.

Table 1. Data splits of FLARE2021.

Data Split	Center	Phase	# Num.
Training (361 cases)	The National Institutes of Health Clinical Center	portal venous phase	80
	Memorial Sloan Kettering Cancer Center	portal venous phase	281
Validation (50 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
Testing (100 cases)	Memorial Sloan Kettering Cancer Center	portal venous phase	5
	University of Minnesota	late arterial phase	25
	7 Medical Centers	various phases	20
	Nanjing University	various phases	50

Table 2. Environments and requirements.

Windows/Ubuntu version	Ubuntu 20.04.2 LTS
CPU	Intel Xeon(R) Silver 4210 CPU @ 2.20GHz
RAM	32×8GB
GPU	NVIDIA Corporation GV100GL [Tesla V100 PCIe 32GB]
CUDA version	11.0
Programming language	Python3.8.8
Deep learning framework	Pytorch (Torch 1.8.1, torchvision 0.9.1)
Specification of dependencies	nnUNet

Table 3. Training protocols.

Data augmentation methods	Rotations, scaling, Gaussian noise, Gaussian blur, brightness, contrast, simulation of low resolution, gamma correction and mirroring.
Initialization of the network	"he" normal initialization
Patch sampling strategy	More than a third of the samples in a batch contain at least one randomly chosen foreground class which is the same as nn-Unet [25].
Batch size	2
Patch size	80×192×160
Total epochs	300
Optimizer	Stochastic gradient descent with nesterov momentum ($\mu = 0.99$)
Initial learning rate	0.01
Learning rate decay schedule	poly learning rate policy: $(1 - epoch/1000)^{0.9}$
Training time	36 hours

5. Results

5.1. Quantitative results for 5-fold cross validation.

The provided results analysis is based on the 5-fold cross validation results and validation cases.

Table 4 illustrates the results of 5-fold cross validation. Figure 4 is the corresponding violin plots of the organ segmentation performance. While high DSC and NSD scores are obtained for liver, kidney and spleen, DSC and NSD scores for pancreas indicating unsatisfactory performance. The violin plot shown in Figure 4 confirms this funding.

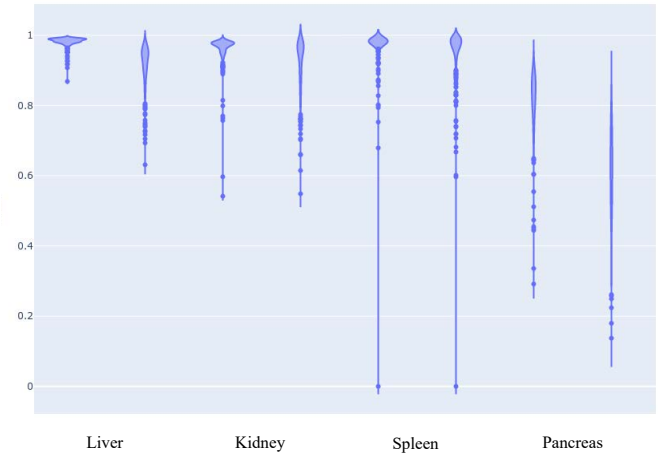

Figure 4. Violin plots of the organ segmentation results (DSC and NSD) of the 5-fold cross validation.

6. Discussion and Conclusion

In this paper, we proposed a lightweight TriSegNet for abdominal organ segmentation. TriSegNet can encode enough spatial information by preserving the resolution of the input CT and capture enough receptive field to obtain high accurate segmentation. The proposed TriSegNet contains three paths. The Spatial Path can preserve the spatial information from the input CT. The Context Path can efficiently obtain a large enough receptive field. The Localization Path utilizes long skip connections to recover the

Table 4. Quantitative results of 5-fold cross validation in terms of DSC and NSD.

Training	Liver		Kidney		Spleen		Pancreas	
	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)	DSC (%)	NSD (%)
Fold-0	98.3±0.8	92.0±4.8	95.9±4.2	91.6±8.4	97.5±3.3	95.6±5.8	81.1±10.4	59.7±14.5
Fold-1	98.2±1.6	92.0±6.1	96.7±3.1	93.0±6.7	97.6±3.2	95.8±6.8	83.9±4.9	62.7±11.7
Fold-2	98.3±1.2	92.0±5.7	96.5±4.6	93.0±6.0	97.9±1.7	96.6±3.9	81.7±9.4	60.5±12.8
Fold-3	98.1±1.2	91.3±5.2	95.7±5.6	91.6±7.5	95.9±11.9	93.7±12.8	80.2±9.1	58.3±13.7
Fold-4	98.4±1.2	93.5±4.4	97.1±1.7	93.8±5.7	97.2±4.1	95.1±6.7	81.3±7.2	61.7±10.9
Average	98.2±1.3	91.7±5.5	96.2±4.1	92.3±7.5	97.2±6.1	95.4±7.8	81.6±8.5	60.6±12.8

original input size and restore the organ details. Compared with other state-of-the-art CNN-based methods, TriSegNet achieves competitive performance with a small size of 9MB. In the evaluation step, the proposed method achieves a mean liver, kidney, spleen and pancreas dice scores of 0.929, 0.752, 0.828 and 0.652, respectively.

Acknowledgment

The authors of this paper declare that the segmentation method they implemented for participation in the FLARE challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers.

References

[1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 1

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99. 1

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440. 2

[4] R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. T. Mok, W. C. W. Chu, D. Wang, and L. Shi, "Automatic Segmentation of Acute Ischemic Stroke From DWI Using 3-D Fully Convolutional DenseNets," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2149–2160, Sep. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8328863/ 2

[5] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, Feb. 2017. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1361841516301839 2

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241. 2

[7] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation," in *European Conference on Computer Vision*. Springer, 2018, pp. 334–349. 2

[8] Z. Zhang, X. Zhang, C. Peng, D. Cheng, and J. Sun, "ExFuse: Enhancing Feature Fusion for Semantic Segmentation," *arXiv preprint arXiv:1804.03821*, 2018. 2

[9] S. Xie, C. Sun, J. Huang, Z. Tu, and K. Murphy, "Rethinking Spatiotemporal Feature Learning For Video Understanding," *arXiv:1712.04851 [cs]*, Dec. 2017, arXiv: 1712.04851. [Online]. Available: http://arxiv.org/abs/1712.04851 2

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9. 3

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 3

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv:1602.07261 [cs]*, Feb. 2016, arXiv: 1602.07261. [Online]. Available: http://arxiv.org/abs/1602.07261 3

[13] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3146–3154. 3

[14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018. 3

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017. 3

[16] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters—improve semantic segmentation by global convolutional network," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017, pp. 1743–1751. 3

[17] L. Sifre and S. Mallat, "Rigid-motion scattering for image classification," PhD Thesis, Citeseer, 2014. 3

[18] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv:1704.04861 [cs]*, Apr. 2017, arXiv: 1704.04861. [Online]. Available: http://arxiv.org/abs/1704.04861 3

[19] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *arXiv:1610.02357 [cs]*, Oct. 2016, arXiv: 1610.02357. [Online]. Available: http://arxiv.org/abs/1610.02357 3

[20] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. Kurfess, "3d separable convolutional neural network for dynamic hand gesture recognition," *Neurocomputing*, vol. 318, pp. 151–161, 2018. 3

[21] R. Ye, F. Liu, and L. Zhang, "3d Depthwise Convolution: Reducing Model Parameters in 3d Vision Tasks," *arXiv:1808.01556 [cs]*, Aug. 2018, arXiv: 1808.01556. [Online]. Available: http://arxiv.org/abs/1808.01556 3

[22] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," *arXiv:1607.08022 [cs]*, Jul. 2016, arXiv: 1607.08022. [Online]. Available: http://arxiv.org/abs/1607.08022 3

[23] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, 2013, p. 3. 4

[24] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[25] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021. 4, 5

[26] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," *Medical Image Analysis*, vol. 71, p. 102035, 2021. 4

[27] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019. 4

[28] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019. 4

[29] H. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. Summers, "Data from pancreas-ct. the cancer imaging archive (2016)." 4

[30] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 556–564. 4

[31] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013. 4

[32] N. Heller, F. Isensee, K. H. Maier-Hein, X. Hou, C. Xie, F. Li, Y. Nan, G. Mu, Z. Lin, M. Han *et al.*, "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge," *Medical Image Analysis*, vol. 67, p. 101821, 2021. 4

[33] N. Heller, S. McSweeney, M. T. Peterson, S. Peterson, J. Rickman, B. Stai, R. Tejpaul, M. Oestreich, P. Blake, J. Rosenberg *et al.*, "An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging." *American Society of Clinical Oncology*, vol. 38, no. 6, pp. 626–626, 2020. 4

[34] J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" *arXiv preprint arXiv:2010.14808*, 2021. 4