3D Deep Learning for Anatomical Structure Segmentation in Multiple Imaging Modalities

Villarini, B., Asaturyan, H., Kurugol, S., Afacan, O., Bell, J.D. and Thomas, E.L.

3D Deep Learning for Anatomical Structure Segmentation in Multiple Imaging Modalities*

Barbara Villarini[1]
Hykoush Asaturyan[2]
School of Computer Science
University of Westminster
London, United Kingdom
[1]b.villarini@westminster.ac.uk
[2]h.asaturyan@my.westminster.ac.uk

Sila Kurugol[3]
Onur Afacan[4]
Department of Radiology
Boston Children's Hospital
& Harvard Medical School
Boston, Massachusetts, USA
[3]Sila.Kurugol@childrens.harvard.edu
[4]Onur.Afacan@childrens.harvard.edu

Jimmy D. Bell[5]
E. Louise Thomas[6]
School of Life Sciences
University of Westminster
London, United Kingdom
[5]j.bell@westminster.ac.uk
[6]l.thomas3@westminster.ac.uk

Abstract—Accurate, automated quantitative segmentation of anatomical structures in radiological scans, such as Magnetic Resonance Imaging (MRI) and Computer Tomography (CT), can produce significant biomarkers and can be integrated into computer-aided diagnosis (CADx) systems to support the interpretation of medical images from multi-protocol scanners. However, there are serious challenges towards developing robust automated segmentation techniques, including high variations in anatomical structure and size, varying image spatial resolutions resulting from different scanner protocols, and the presence of blurring artefacts. This paper presents a novel computing approach for automated organ and muscle segmentation in medical images from multiple modalities by harnessing the advantages of deep learning techniques in a two-part process. (1) a 3D encoder-decoder, Rb-UNet, builds a localisation model and a 3D Tiramisu network generates a boundary-preserving segmentation model for each target structure; (2) the fully trained Rb-UNet predicts a 3D bounding box encapsulating the target structure of interest, after which the fully trained Tiramisu model performs segmentation to reveal organ or muscle boundaries for every protrusion and indentation. The proposed approach is evaluated on six different datasets, including MRI, Dynamic Contrast Enhanced (DCE) MRI and CT scans targeting the pancreas, liver, kidneys and iliopsoas muscles. We achieve quantitative measures of mean Dice similarity coefficient (DSC) that surpasses or are comparable with the state-of-the-art and demonstrate statistical stability. A qualitative evaluation performed by two independent experts in radiology and radiography verified the preservation of detailed organ and muscle boundaries.

Index Terms—automated organ segmentation, 3D deep learning, CADx system, anatomical structure, multiple modalities

I. INTRODUCTION

Medical imaging using non-invasive techniques has rapidly evolved in the last decade, providing detailed and more reliable images of anatomy in the human body [1]. Every year millions of abdominal radiological scans of Magnetic Resonance Imaging (MRI) and Computer Tomography (CT) modality are acquired. The accurate analysis and segmentation of these scans' anatomical structures can produce significant biomarkers to examine a medical condition and provide additional guidance towards subject stratification after a diagnosis or before a clinical trial [2]. Although expert-led, manual segmentation in images from radiological scans can produce clinically acceptable results for analysis, it is time-consuming, sometimes prone to inter-observer variability, but above all, it is challenging to replicate on a large-scale of multiple thousands of scans. On the other hand, recent advances in CADx have demonstrated the vital role of automated segmentation in raising biomedical research quality concerning abdominal organs [3]. Nonetheless, there are serious challenges in developing robust automated quantitative segmentation techniques, mostly but not limited to high variations in organ structure and size, varying spatial resolutions, and the imaged quality of radiological scans of interest acquired from different scanner modalities and protocols. Despite CT scanning taking less time and costing less than MRI scanning, some types of cancers or lesions, such as certain liver cancers, can be harder to identify where otherwise detailed in an MRI scan. On the other hand, MR imaging may not always differentiate between excessive fluid edema and cancerous tissue due to similar greyscale intensity and structure. Furthermore, both MRI and CT scanning can produce blurring motion-based artefacts resulting from breathing, cardiac movement and blood flow.

In recent research literature, various automated organ segmentation methods have been proposed in the scheme of multi-atlas label propagation (MALP) [4], [5] and convolutional neural networks [6]–[8]. MALP segmentation methods employ "atlases", which are labelled regions of interest (e.g. organs) and corresponding intensities in radiological image volumes. Such methods usually employ image registration to align the atlases to an (unseen) test image volume, which are then combined using label fusion to determine the overall segmentation [9]. Another set of approaches that often but not always integrate into MALP is statistical shape modelling [10]. Such methods define a template shape for a given structure

representing the organ of interest with reliance upon control points along the boundary, after which the entire shape is deformed to match the test image volume. The approach reported in [11] performs multi-organ segmentation by combining spatial interrelations with multiple probabilistic atlases and incorporating prior knowledge into the model using shape representations of multiple organs. MALP methods in recent years have achieved high accuracies in mean Dice similarity coefficient (DSC) ($\geq 70\%$). However, inter-patient registration is computationally expensive and extremely poor to imaged organs that possess high variability such as the pancreas [12].

In the last decade, the rise of convolutional neural networks (CNNs) has boosted the performance of several imaging tasks using large-scale data for semantic segmentation [13] and has successfully applied to medical image segmentation tasks, especially for abdominal organs that are highly deformable and possess vague edge boundaries. Unlike MALP techniques, CNNs do not require selecting a specific atlas nor require deformable registration from training datasets to a target image. The publication [14] presents a deep learning segmentation architecture, DenseVNet, for eight different organs relevant for navigation in endoscopic pancreatic and biliary procedures, including the pancreas. The DenseVNet segmentation network aims to enable high-resolution activation maps through (memory-efficient) dropout layers and reusing feature maps. Two abdominal CT datasets containing a total of 90 image volumes were employed to evaluate this approach. The authors of [15] extend organ segmentation in MRI to present a method that combines a MALP with CNN. This approach builds on previous work described in [16] by incorporating weighting schemes to support class imbalance and a specialised organ region-of-interest selection. Later, spatial information from multi-atlases and CNNs are optimised jointly and applied for organ segmentation. This method is evaluated on a dataset containing 48 whole MRI body volumes. The CNN methods described above yield DSC results that outperform the reported MALP approaches and produce a lower standard deviation. However, such CNNs are prone to suffer from an imbalance between classes and overfitting during the network training stage [17], ignoring features related to the organ of interest during the testing stage. Furthermore, many segmentation methods have been performed on CT modality and do not address the additional difficulties of image artefacts and higher blurred boundaries between organs in MRI.

Considering the above challenges, the contributions of this paper are: (a) a novel and robust automated 3D deep learning approach for automated quantitative organ and muscle segmentation in volumes from different modalities; (b) this approach employs volumetric information instead of 2D feature learning and is modular, scalable and generalisable; (c) the proposed approach is evaluated on six different datasets of MRI, Dynamic Contrast Enhanced (DCE) MRI and CT modality, targeting four distinct abdominal structures (pancreas, liver, kidneys and iliopsoas muscles); (d) we achieve mean DSC scores that outperform or are comparable with state-of-the-art and demonstrate high statistical stability, given the differences

in image quality. Section II explains the methodology of the automated segmentation approach. Section III presents and discusses the quantitative and qualitative results. Section IV provides a conclusion for the proposed approach, including reference to future work in the area of multi-organ segmentation.

II. METHODOLOGY

In 2D deep learning, the 3D radiological scans or image volumes are processed slice by slice (2D image), whereas 3D deep learning employs volumetric information instead of pixel information in 2D. The proposed approach has a two-part process: the first part develops a localisation model known as 3D Rb-UNet to embody the organ of interest, and the second part performs a detailed organ segmentation through a 3D Tiramisu network. The testing stage process an original radiological volume to predict the minimal bounding box that captures the organ of interest and then processes the cropped image volume to predict the target organ's labels. The training stage and testing stage for each part are shown in Fig. 1.

A. Training stage

a) Detection and Localisation: The first part of the training stage aims to localise the organ of interest and develops a model defined as 3D Rb-UNet. In this model, residual connections introduced in [18] are added at each block of a baseline 3D U-Net [19] architecture, connecting the input of convolutional layers at each scale to the outputs of the corresponding layer. This architecture has the advantage of alleviating the vanishing gradient problem. To begin, the size of the image volume is reduced to $64 \times 64 \times 64$ to limit computational costs and have sufficient resolution necessary for localisation and near isotropic resolution across x, y and z dimensions. Inspired by the technique described in [20], for datasets that contain 4D image volumes, the temporal dimension is reduced from 150 seconds (i.e. 150 image volumes) to 5 seconds (i.e. 5 image volumes) using Principal Component Analysis (PCA). In this way, performance improves while avoiding memory overload. Data augmentation is generated as image scaled translations in the range [1, 4] pixels. The weighted cross entropy loss function denoted as W_{CE} is employed to compensate for the class imbalance presented by diverse anatomical structures:

$$W_{CE} = -\frac{1}{N} \sum_{i=1}^{N} w_i^c \left[\hat{p}_i \log p_i + (1 - \hat{p}_i) \log(1 - p_i) \right] \quad (1)$$

where N is number of voxels in an image volume, p_i is the probability of voxel i belonging to the foreground in each output channel and $\hat{p}_i$ represents the true ground-truth label in the corresponding input channel. The w_i^c is fixed as inversely proportional to the probability of voxel i belonging to the foreground class. Afterwards, softmax with weighted cross-entropy loss is used to compare the network's output with the corresponding ground-truth.
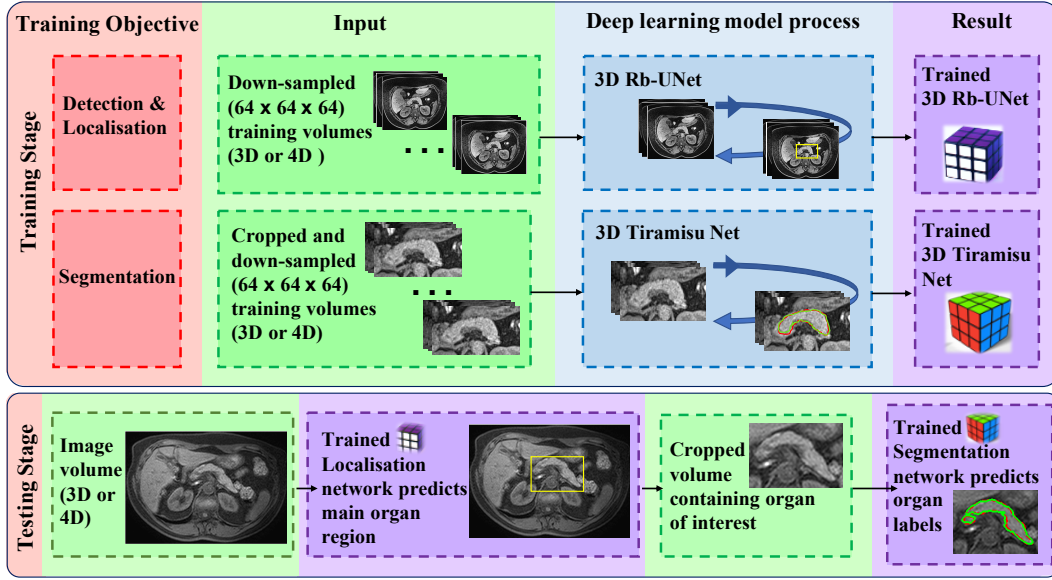
Fig. 1. Overview of the proposed automated organ segmentation approach. The training stage simultaneously develops a network (3D Rb-UNet) for localising the organ, and a segmentation network (3D Tiramisu) to predict the labels that correspond to the organ of interest. The testing stage processes an original scan (3D or 4D volume) to predict the bounding box capturing the organ and then processes the cropped image volume to predict the labels of that organ.

b) Segmentation: The second part of the training stage develops a 3D Tiramisu model [21] using a uniformly cropped region where the organ of interest is fully present, discarding unrelated background information. The target organ's main region is captured using a minimal bounding box generated via corresponding ground-truth labels, after which the image volume dimensions are reduced to $64 \times 64 \times 64$. This input (with the temporal dimension reduced to 5 if using 4D data) feeds into the 3D Tiramisu network for training the segmentation model using the same weighted cross-entropy loss as in (1). This model's prediction aims to classify a voxel as representing the target "organ" or "non-organ". The Tiramisu model builds upon Densely Connected Convolutional Networks (DenseNets) [22]. This model builds upon DenseNets to work as Fully Convolutional Nets (FCN) by adding an upsampling path to compensate for the input's full resolution. However, building an upsampling path would result in an exponential rise in features maps. Thus, Tiramisu mitigates this problem by implementing skip connections where the feature maps from each step in the downsampling path are concatenated with feature maps from the corresponding step in the upsampling path. This allows the recovery of fine-grained information from the downsampling layer. This architecture builds a very deep FCN DenseNet while limiting the number of parameters. Architecture details are provided in Fig. 2.

The bypass with residual identity connections for convolutional blocks at each scale in 3D Rb-UNet improves convergence and allows faster training, which is excellent for localisation. Meanwhile, the 3D Tiramisu provides a higher capacity with multi-layer feature concatenation; it delivers very detailed boundary-preserving fine segmentation given a localised organ of interest as the primary input.

Layer
Batch Normalization
ReLu
3 x 3 Convolution
Dropout = 0.2

Transition Down (TD)
Batch Normalization
ReLu
1 x 1 Convolution
Dropout = 0.2
2 x 2 Max Pooling

Transition up (TU)
3 x 3 Transposed convolution Stride = 2

Architecture
Input m = 3
DB (4 Layers) + TD, m = 112
DB (5 Layers) + TD, m = 192
DB (7 Layers) + TD, m = 304
DB (10 Layers) + TD, m = 464
DB (12 Layers) + TD, m = 656
DB (15 Layers) , m = 896
TU + DB (12 Layers), m = 1088
TU + DB (10 Layers), m = 816
TU + DB (7 Layers), m = 578
TU + DB (5 Layers), m = 384
TU + DB (4 Layers), m = 256
1 x 1 Convolution, m = 2 (or 3)
Softmax

Fig. 2. Architecture details of Tiramisu network. m corresponds to the total number of feature maps at the end of a block.

B. Testing stage

Initially, the fully trained 3D Rb-UNet performs a coarse segmentation, i.e. voxel-based prediction using an unseen (test) image volume. For target structures of interest consisting of two distinct parts (e.g. left and right kidney or left and right psoas muscle), there are three classes at this stage: "left", "right" and "background". Otherwise, there are two main classes of "foreground" and "background" (e.g. "pancreas" versus "non-pancreas" or "liver" versus "non-liver"). Afterwards, the image volume resamples to its original size and one or two bounding boxes are generated to "crop out" the organ of interest throughout the entire 3D or 4D volume. The

cropped test volume is fitted to $64 \times 64 \times 64$ dimensionality and processed through the fully trained 3D Tiramisu model, which performs detailed voxel-wise predictions on whether each voxel corresponds to the organ of interest (foreground) or otherwise (background). Afterwards, the predicted organ binary mask is resampled to its original size and inserted into the primary input image volume's corresponding spatial position.

C. Data and Experimental Setup

The proposed approach is evaluated on the following datasets, each containing D image volumes.

- Pancreas: $D = 216$ (MRI-A). 196/20 for train/test. Annotated on 3D fat suppressed T2-weighted MRI obtained using a Philips Intera 1.5 Tesla (T) scanner (50 axial slices of 384×384, voxel size $0.9766 \times 0.9766 \times 2$ mm).
- Pancreas: $D = 132$ (MRI-B). 112/20 for train/test. Annotated on 3D fat suppressed T2-weighted MRI obtained using a Siemens Trio 3T scanner (80 axial slices of 320×260, voxel size $1.1875 \times 1.1875 \times 1.6$ mm).
- Pancreas: $D = 82$ (CT-NIH). 62/20 for train/test. Annotated on 3D contrast-enhanced CT acquired on Philips and Siemens MDCT scanners (161-466 axial slices of 512×512, voxel sizes range from 1.5-2.5 mm in x-y-z). Publicly available at http://dx.doi.org/10.7937/K9/TCIA.2016.tNB1kqBU.
- Liver: $D = 30$ and 20/10 for train/test. Annotated on 3D T2-weighted MRI obtained using a Siemens Trio 3T scanner (370 axial slices of 224×174, voxel size $2.2321 \times 2.2321 \times 3$ mm).
- Iliopsoas muscles: $D = 30$ and 20/10 for train/test. Annotated on 3D T2-weighted MRI obtained using a Siemens Trio 3T scanner (370 axial slices of 224×174, voxel size $2.2321 \times 2.2321 \times 3$ mm).
- Kidneys: $D = 60$ and 34/26 for train/test. Annotated on 4D DCE-MRI scans of acquired at 3 Tesla (T) for 6 minutes after injecting Gadavist using a motion-robust, radial stack-of-stars 3D FLASH sequence (32 coronal slices of 224×224, voxel size $1.25 \times 1.25 \times 3$ mm).

The proposed method has been implemented using Python 3.0 and Keras in an i7-5930K-CPU at 3.5 GHz (NVIDIA GeForce TitanX). The optimisation algorithm used for training is Adam [23] with an initial learning rate of 0.0001. The hyperparameters include a reduction rate (0.8), growth rate (12), momentum (0.9), weight decay (10^{-8}) and a drop-out rate (0.2). The learning rate drop period is 50 and the learning rate drop factor is 0.5. The maximum number of epochs is 400 and the size of the mini-batch to use for each training iteration is set to 4 (with a validation split of 0.5).

D. Evaluation

A popular metric used to evaluate the segmentation accuracy is the Dice similarity coefficient (DSC), which is defined as $DSC = 2(|G \cap S|)/(|G| + |S|)$, where G is a volumetric ground-truth and S is the corresponding automated segmentation labels. A second commonly used metric, Jaccard index

(JI), is considered and defined as $JI = (|G \cap S|)/(|G \cup S|)$. Both metrics can be described as the quotient of similarity between 0 and 1 (or 0 to 100%).

TABLE I
DSC AND JI FOR AUTOMATED PANCREAS SEGMENTATION IN MRI

Method	MRI-A		MRI-B	
	DSC(%)	JI (%)	DSC (%)	JI (%)
2D UNet [19]	69.1±10.2	53.8±14.2	72.8±7.5	67.9±10.2
2D FCN [13]	70.2±8.5	63.5±13.5	70.9±7.7	65.4±13.5
Deeporgan [24]	44.5±25.2	32.7±29.4	50.1±22.7	44.9±12.0
Multiorgan [14]	52.6±17.1	44.1±20.7	55.8±18.6	49.9±18.7
Casc. 3D FCN [25]	65.2±10.1	52.2±15.3	69.6±11.5	61.2±15.9
Geo. descript. [26]	79.6±5.7	66.5±7.9	81.6±5.1	69.2±7.1
Hausdorff-Sine [27]	84.1±4.6	72.9±6.5	85.7±2.3	75.1±3.5
Proposed	**89.9±3.4**	**81.9± 5.6**	**90.2±5.1**	**82.6±7.8**

TABLE II
DSC AND JI FOR AUTOMATED PANCREAS SEGMENTATION IN CT

Method	CT-NIH	
	DSC (%)	JI (%)
2D UNet [19]	79.7±7.6	66.3±4.0
2D FCN [13]	80.3±9.0	67.1±4.7
Recurrent NN [16]	82.4±6.7	70.6±9.0
Holistically-nested CNN [12]	81.3±6.2	68.5±3.2
Cascade 3D FCN [25]	76.8±9.4	62.3±4.9
Geo. Descriptors [26]	79.3±4.4	66.1±6.2
Hausdorff-Sine [27]	83.1±5.3	71.4±7.4
Proposed	**84.7±7.9**	**74.2±11.4**

III. RESULTS AND DISCUSSION

The proposed approach is applied to segment the pancreas using two MRI datasets (MRI-A and MRI-B) and one CT dataset (CT-NIH). It is noted that MRI-A and MRI-B were produced from subjects that showed early signs of type 2 diabetes, while CT-NIH was produced from healthy kidney donors scanned before nephrectomy. This approach is directly evaluated against the state-of-the-art segmentation models as shown in Table I and Table II, achieving a mean DSC (%) $\pm$ Standard deviation (SD) of 89.9 $\pm$ 3.4 for MRI-A, 90.2 $\pm$ 5.1 for MRI-B and 84.7 $\pm$ 7.9 for CT scans. The proposed approach outperforms other works in terms of mean DSC and JI for all three datasets. The proposed approach is extended to a broader spectrum of MRI sequences and organs. As shown in Table III, the segmentation results for the liver achieves a mean DSC of 95.64 $\pm$ 1.31% and mean JI of 91.66 $\pm$ 2.42% reflecting the ability to better address a diversity of artefact-prone imaging compared with the state-of-the-art, which achieved mean DSC scores of 94% [28], 72.9% [19] and 94.5% [29]. The segmentation results for iliopsoas muscles deliver a mean DSC of 88.41 $\pm$ 2.39% and a mean JI of 78.56 $\pm$ 3.59%, raising the state-of-the-art when compared with 72.3% in mean JI [30].

The kidneys dataset contains pediatric DCE-MRI 4D volumes with a clinical "normal" assessment and varying levels of hydronephrosis that exhibit irregular shape, size and location. Evaluating the "normal" dataset (30 volumes) achieves a mean DSC of 90.48% $\pm$ 1.56% and a mean JI of 82.64 $\pm$ 2.61%

| Anatomical Structure	Clinical Assessment	Modality	Data dimension	Dataset Size	DSC (%)
Pancreas	Early type 2 diabetes	MRI	3D	216	89.9 ± 3.4
Pancreas	Early type 2 diabetes	MRI	3D	132	90.2 ± 5.1
Pancreas	Normal	CT	3D	82	84.7 ± 7.9
Liver	Normal	MRI	3D	30	95.64 ± 1.31
Iliopsoas muscles	Normal	MRI	3D	30	88.41 ± 2.39
Kidneys	Normal	DCE-MRI	4D	30	90.48 ± 1.56
Kidneys	Hydronephrosis	DCE-MRI	4D	30	86.44 ± 3.84
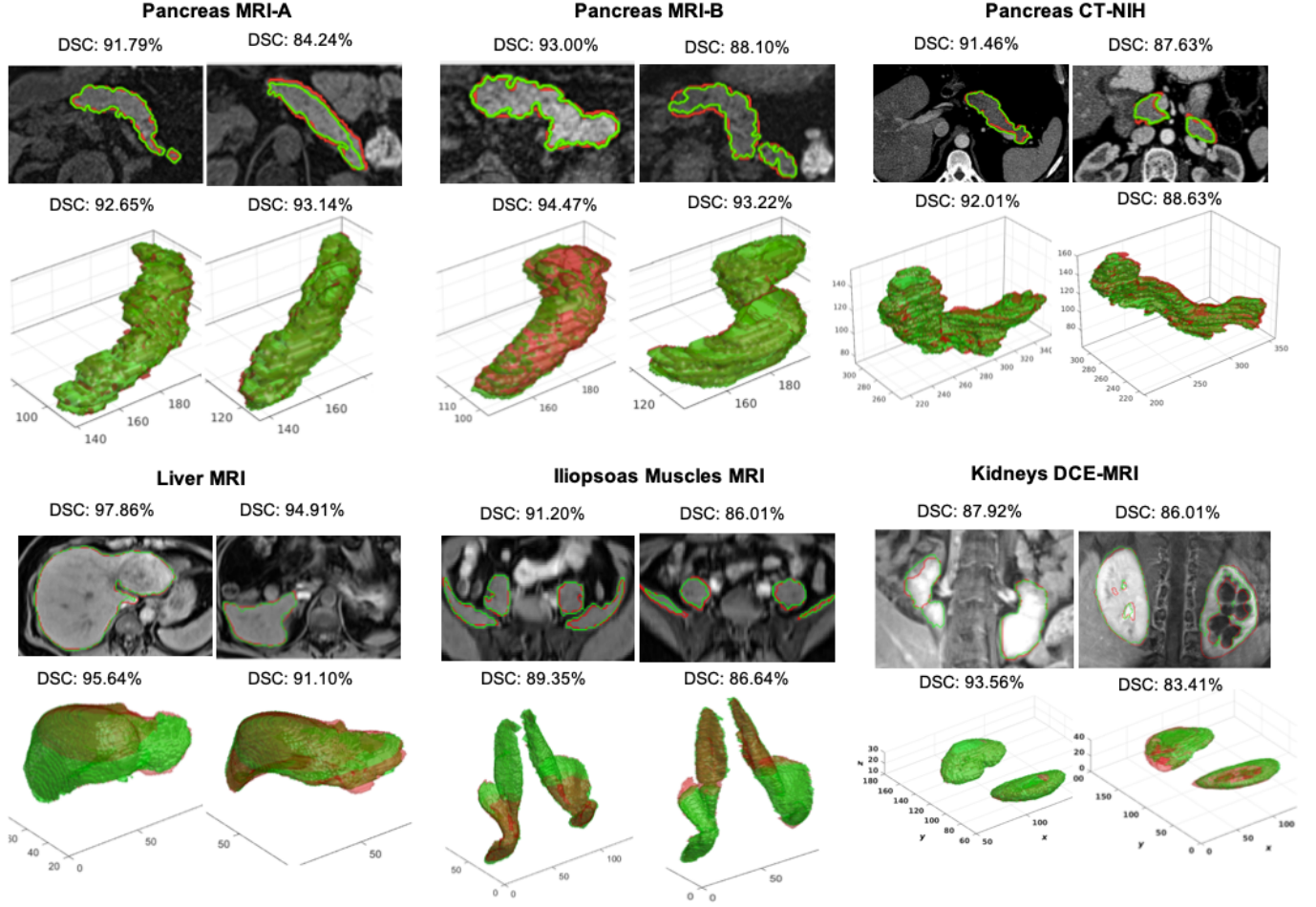
Fig. 3. For each dataset, two different medical scans are shown with superimposed segmentation outcome (green) and ground-truth (red), and DSC on top.

as shown in Table III. The low standard deviation demonstrates robustness toward intensity variation of contextual information. In contrast, the 3D U-Net (including the Rb-UNet localisation) delivers poorer quantitative accuracy, as in DSC of 82.89 ± 5.0% indicating weak stability. Although 3D DenseNet (including the Rb-UNet localisation) delivers better accuracy than 3D U-Net, the approach is still inferior to the proposed approach, achieving a mean DSC of 85.09 ± 6.00%.

The segmentation results obtained for the "hydronephrosis" dataset (30 volumes) continues to demonstrate statistical sta-bility, achieving a mean DSC of 86.44 ± 3.84% and mean JI of 76.29 ± 5.93%. The baseline 3D U-Net suffers from high instability, trailing behind with a mean DSC of 78.94 ± 9.10%, while the performance of the 3D DenseNet produces a mean DSC of 84.51 ± 3.30%. Fig. 3 displays the results for two image volumes from each of the six datasets. For each image volume, the top row shows a 2D slice with the segmentation outcome (green) and ground-truth (red) superimposed and the DSC. Similarly, the bottom row shows the 3D reconstruction of the entire target anatomical structure.

Qualitative feedback from two independent experts in radiology and radiography confirmed that the approach produces detailed organ and muscle contouring for every protrusion and indentation, as opposed to an approximate tracing, which is an essential determinant for stratifying between "normal" and "abnormal" variations in a clinical assessment.

IV. CONCLUSION

There are serious challenges towards developing robust segmentation methods, mostly but not limited to high variations in anatomical structure and size and distinct datasets that correspond to different scanner modalities, protocols and image resolution. The proposed approach generates automated, accurate organ or muscle segmentation via 3D deep learning instead of 2D feature learning by exploiting volumetric contextual information to perform localisation followed by fine segmentation of the target anatomical structure. The proposed approach is applied to four distinct abdominal structures of high intervariability. We achieve robust segmentation performance better than or close to the state-of-the-art using CT, MRI and DCE-MRI and reports higher statistical stability with lower standard deviation measures. The segmentation approach can extend to serve as a tool for the classification of clinical measures and could provide an indication or prognostic biomarker about the progression or severity of a medical condition.

REFERENCES

[1] J. Lohrke, T. Frenzel, J. Endrikat, F. C. Alves, T. M. Grist, M. Law, J. M. Lee, T. Leiner, K.-C. Li, K. Nikolaou, M. R. Prince, H. H. Schild, J. C. Weinreb, K. Yoshikawa, and H. Pietsch, "25 years of contrast-enhanced mri: Developments, current challenges and future perspectives," *Adv Ther.*, vol. 33, no. 1, pp. 1–28, 2016.

[2] H. Asaturyan, E. L. Thomas, J. D. Bell, and B. Villarini, "A framework for automatic morphological feature extraction and analysis of abdominal organs in mri volumes," *J. Med. Syst.*, vol. 43, no. 12, p. 334, 2019.

[3] C.-M. Chen, Y.-H. Chou, N. Tagawa, and Y. Do, "Computer-aided detection and diagnosis in medical imaging," *Comput Math Methods Med*, vol. 2013, pp. 556–564, 2013.

[4] T. Tong, R. Wolz, Z. Wang, Q. Gao, K. Misawa, M. Fujiwara, K. Mori, J. V. Hajnal, and D. Rueckert, "Discriminative dictionary learning for abdominal multi-organ segmentation," *Med Image Anal*, vol. 23, no. 1, pp. 92–04, 2015.

[5] K. Karasawa, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, C. Chengwen, G. Zheng, D. Rueckert, and K. Mori, "Multi-atlas pancreas segmentation: Atlas selection based on vessel structure," *Med Image Anal*, vol. 39, pp. 18–28, 2017.

[6] Y. Wang, Y. Zhou, P. Tang, W. Shen, E. K. Fishman, and A. L. Yuille, "Training multi-organ segmentation networks with sample selection by relaxed upper confident bound," *arXiv preprint arXiv:1804.02595*, 2018.

[7] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.

[8] B. Kayalibay, G. Jensen, and P. van der Smagt, "Cnn-based segmentation of medical imaging data," *arXiv preprint arXiv:1701.03056*, 2017.

[9] P. A. Yushkevich, H. Wang, J. Pluta, and B. B. Avants, "From label fusion to correspondence fusion: a new approach to unbiased groupwise registration," in *IEEE on CVPR*, pp. 956–963, IEEE, 2012.

[10] I. L. Dryden and K. V. Mardia, *Statistical shape analysis: with applications in R*. John Wiley & Sons, 2016.

[11] T. Okada, M. G. Linguraru, M. Yoshida, M. Hori, R. M. Summers, Y.-W. Chen, N. Tomiyama, and Y. Sato, "Abdominal multi-organ segmentation of ct images based on hierarchical spatial modeling of organ interrelations," in *Abdom Radiol*, pp. 173–180, Springer Berlin Heidelberg, 2012.

[12] H. Roth, A. Farag, L. Lu, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *MICCAI*, pp. 451–450, 2016a.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE CVPR*, pp. 3431–3440, 2015.

[14] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, "Automatic multi-organ segmentation on abdominal CT with dense V-Networks," *IEEE Trans Med Imaging*, 2018.

[15] V. V. Valindria, I. Lavdas, J. Cerrolaza, E. O. Aboagye, A. G. Rockall, D. Rueckert, and B. Glocker, "Small organ segmentation in whole-body mri using a two-stage fcn and weighting schemes," in *MLMI*, pp. 346–354, Springer, 2018.

[16] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," *CoRR*, vol. abs/1707.04912, 2017.

[17] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Fourth Int. Conf. 3DV*, pp. 565–571, 2016.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 770–778, 2016.

[19] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, pp. 234–241. Cham: Springer International Publishing, 2015.

[20] M. Haghighi, S. K. Warfield, and S. Kurugol, "Automatic renal segmentation in dce-mri using convolutional neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1534–1537, IEEE, 2018.

[21] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 11–19, 2017.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*, pp. 4700–4708, 2017.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[24] H. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. Turkbey, and R. Summers, *DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation*, pp. 556–564. Springer International Publishing, 2015a.

[25] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori, "An application of cascaded 3d fully convolutional networks for medical image segmentation," *Comput Med Imaging Graph*, vol. 66, pp. 90–99, 2018.

[26] H. Asaturyan, A. Gligorievski, and B. Villarini, "Morphological and multi-level geometrical descriptor analysis in ct and mri volumes for automatic pancreas segmentation," *Comput Med Imaging Graph*, vol. 75, pp. 1–13, 2019.

[27] H. Asaturyan, E. L. Thomas, J. Fitzpatrick, J. D. Bell, and B. Villarini, "Advancing pancreas segmentation in multi-protocol mri volumes using hausdorff-sine loss function," in *MLMI*, pp. 27–35, Springer, 2019.

[28] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D. Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze, "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," 2017.

[29] C. Li, X. Wang, S. Eberl, M. Fulham, Y. Yin, J. Chen, and D. D. Feng, "A likelihood and local constraint level set model for liver tumor segmentation from ct volumes," *IEEE Trans. Biomed.*, vol. 60, no. 10, pp. 2967–2977, 2013.

[30] N. Kamiya, X. Zhou, H. Chen, C. Muramatsu, T. Hara, R. Yokoyama, M. Kanematsu, H. Hoshi, and H. Fujita, "Automated segmentation of psoas major muscle in x-ray ct images by use of a shape model: preliminary study," *Radio. Phys. and Tech.*, vol. 5, no. 1, pp. 5–14, 2012.