Evaluation of U-Net CNN Approaches for Human Neck MRI Segmentation

Abdulla Al Suman[1], Yash Khemchandani[2], Md. Asikuzzaman[1], Alexandra Louise Webb[3], Diana M. Perriman[3,4], Murat Tahtali[1], and Mark R. Pickering[1]
[1] School of Engineering and Information Technology
The University of New South Wales, Canberra, Australia
[2] Indian Institute of Technology Bombay, Mumbai, India
[3] Medical School, Australian National University, Canberra, Australia
[4] Trauma and Orthopaedic Research Unit, Canberra Hospital, Canberra, Australia

***Abstract*—The segmentation of neck muscles is useful for the diagnoses and planning of medical interventions for neck pain-related conditions such as whiplash and cervical dystonia. Neck muscles are tightly grouped, have similar appearance to each other and display large anatomical variability between subjects. They also exhibit low contrast with background organs in magnetic resonance (MR) images. These characteristics make the segmentation of neck muscles a challenging task. Due to the significant success of the U-Net architecture for deep learning-based segmentation, numerous versions of this approach have emerged for the task of medical image segmentation. This paper presents an evaluation of 10 U-Net CNN approaches, 6 direct (U-Net, CRF-Unet, A-Unet, MFP-Unet, R2Unet and U-Net++) and 4 modified (R2A-Unet, R2A-Unet++, PMS-Unet and MS-Unet). The modifications are inspired by recent multi-scale and multi-stream techniques for deep learning algorithms. T1 weighted axial MR images of the neck, at the distal end of the C3 vertebrae, from 45 subjects with real-time data augmentation were used in our evaluation of neck muscle segmentation approaches. The analysis of our numerical results indicates that the R2Unet architecture achieves the best accuracy.**

***Index Terms*—U-Net, Deep Learning, Neck muscles, Segmentation, Whiplash.**

I. Introduction

The accurate quantification of a muscle's size (volume and cross-sectional area) and composition (quantity and distribution of adipose tissue) using magnetic resonance imaging (MRI) is important for advancing the diagnosis and management of disease and injury as it enhances the understanding of conditions such as whiplash and cervical dystonia (CD) [1], [2]. However, the manual segmentation of muscles from MRIs is a time-consuming activity influenced by the subjectivity of the researcher [3]. Segmenting neck muscles is particularly challenging because there are more than 20 muscles with complex relationships located in close proximity within a small region [4], [5]. Also, these muscles are similar in composition and exhibit large anatomical variability. The images also contain areas of intra-muscular fat and background clutter of a similar intensity, as shown in Fig. 1. Consequently, identifying muscle boundaries is difficult, especially for deeper and pathological muscles which have poorly visualized muscle-fascial borders [6]. Furthermore, slice-wise manual segmentation is laborious and time-consuming, which limits its application in large studies and clinical settings [3]. The use of automated segmentation methods has been suggested as a solution for achieving accurate and rapid quantification of the sizes and compositions of these challenging muscles.

There are many medical image segmentation methods in the literature, such as template matching, deformable model fitting, edge detection and learning-based approaches. However, deep learning-based techniques have revolutionized segmentation tasks [7]; in particular, the U-Net [8] architecture is popular for biomedical image segmentation and has attained remarkable success due to its flexibility [7]. Over the last few years, there have been rapid advances based on the U-Net architecture for medical and natural image segmentation [9]. The large list of U-Net based approaches requires a comparative study to choose the best algorithms for further research.

In this paper, we will present an evaluation of U-Net-based approaches for neck muscle segmentation in MRIs. The approaches chosen for evaluation include direct and modified models, with the former based on slight variations in the U-Net architecture (no major architectural change) and the latter based on integration of some recent techniques for deep learning-based segmentation.

Deep learning-based segmentation for cell segmentation was recently evaluated by Caicedo *et al.* [10]. In this paper the authors evaluated 5 methods, including deep learning strategies (U-Net, DeepCell), classical machine learning (Random Forest) and classical image processing (advanced and basic CellProfiler), for the segmentation of cell nuclei in fluorescent images. In their experiment, they used 200 images with $23,165$ manually annotated nuclei. The performances of the Random Forest and advanced and basic CellProfiler methods were found to be worse than those of the deep learning strategies. Hansch *et al.* [11] compared 2D U-Net, 2D ensemble U-Net and 3D U-Net for parotid gland segmentation from a head and neck computed tomography (CT) in the 2015 Medical Image Computing and Computer-assisted Intervention (MICCAI) challenge. The 2D ensemble U-Net was actually a combination of three 2D U-Nets arranged in parallel to work on sagittal, coronal and axial patches, respectively, in which predictions of the individual 2D models were combined through a majority voting rule.

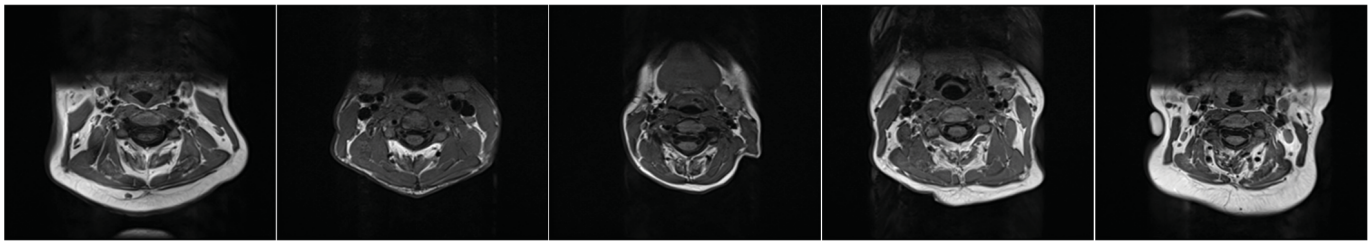

Fig. 1. Anatomical variation in our dataset (image samples from different individuals).

In this paper we will evaluate 10 popular deep learning-based segmentation models, 6 direct (U-Net, CRF-Unet, A-Unet, MFP-Unet, R2Unet and U-Net++) and 4 modified (R2A-Unet, R2A-Unet++, PMS-Unet and MS-Unet). Brief descriptions of the models and an explanation of the modifications incorporated are provided in Section II-B. The code for the models have been made available to the public by the corresponding authors. As our research is ongoing, more models will be investigated in future and we believe that this study will help further research in the medical image segmentation community.

II. Method and Materials

A. Image Acquisition and Pre-processing

Our dataset was collected from an ongoing larger study entitled "Characterizing whiplash injury using magnetic resonance imaging" conducted by Medical School of Australian National University and ethics approval was obtained for this study. The dataset consists of 45 patients (28 females and 17 males) aged from 18 to 36 years, weighing 50 to 100 kilograms and with heights of 1.5 to 1.8 meters. The dataset was extended through real-time data augmentation during training of the models. All the subjects, who were suffering neck pain, provided written informed consent to participate in the experiments, with their personal identifying information removed prior to the research being conducted. The data was obtained in multiple medical centers over a two-year period using 3-Tesla MR scanners (Siemens, Skyra, Erlangen, Germany) using different MRI protocols. Although the images were different in their spatial resolutions, intensity ranges and levels of contrast, they were later normalized. Axial T1-weighted spin echo images with sizes of 256×256 were obtained from between the participants' occiput and T1 vertebral levels (45 slices, each with a repetition time of 746-1140 ms, echo time of 15-16 ms, slice thickness of 4 mm, inter-slice gap of 4 mm, flip angle of 70°, pixel bandwidth of 300 Hz/pixel and percent phase field of view of 100). As the images were obtained using different MRI protocols, the inhomogeneity of the MRI intensity of each was corrected individually by varying the parameters of the multiplicative intrinsic component optimization (MICO) method [12]. One student from the university's medical school, with extensive training on cervical spine anatomy, manually segmented the left and right muscles (sternocleidomastoid, semispinalis capitis and splenius capitis) at the top of the C3 vertebral level using a Matlab graphical user interface (GUI) specially constructed for this study. Later, the annotations were validated by two other anatomical experts from the medical school. The segmentation masks will be shown using six colors for the six muscles (yellow (M1), cyan (M2), red (M3), magenta (M4), blue (M5) and green (M6)) of a patient, as shown in the last column in Fig. 2. The inhomogeneity-corrected and ground-truth images were aligned with a population average using an affine transformation and edge position difference (EPD) similarity measure-based registration technique [13], [14]. The image of the population average was generated using the registration algorithm in [13], [15] and the demons deformable registration algorithm [16]. Fig. 2 shows a raw MR image, its MICO-corrected version, the population average for this study and the ground truth segmentation for one subject.

B. Deep CNNs

Deep learning-based segmentation methods have recently been shown to achieve more promising results than traditional methods in medical image segmentation applications. In particular, the U-Net [8] architecture is a popular approach due to its good performance. Many researchers have tried to improve its performance by integrating additional techniques into its original architecture. In our experiments, we evaluated some recent prominent U-Net-based networks, including the U-Net [8], U-Net with a conditional random field (CRF-Unet) [17], attention U-Net (A-Unet) [18], nested U-Net or U-Net++ [19], multi-feature pyramid (MFP)-Unet [20], R2Unet [21]. We also evaluated networks with more comprehensive modifications, including the multi-scale U-Net (MS-Unet), parallel multi-scale U-Net (PMS-Unet), recurrent residual attention U-Net (R2A-Unet), and R2A-Unet++, in our neck muscles segmentation research.

The A-Unet was included in our evaluation study because its attention gate can automatically increase any model's prediction accuracy and sensitivity by concentrating on target structures and neglecting irrelevant regions. The A-Unet showed improved accuracy over that of the U-Net for two 3D CT abdominal datasets [18]. The CRF-Unet uses CRFs as recurrent neural networks [22]. A CRF establishes spatial constraints among labels to reduce false labeling which occurs due to local minima in training and image noise, and increases segmentation accuracy in natural and medical images [7], [17]. The MFP-Unet uses a feature pyramid to extract feature maps from all the blocks of an expanding path in a semantic

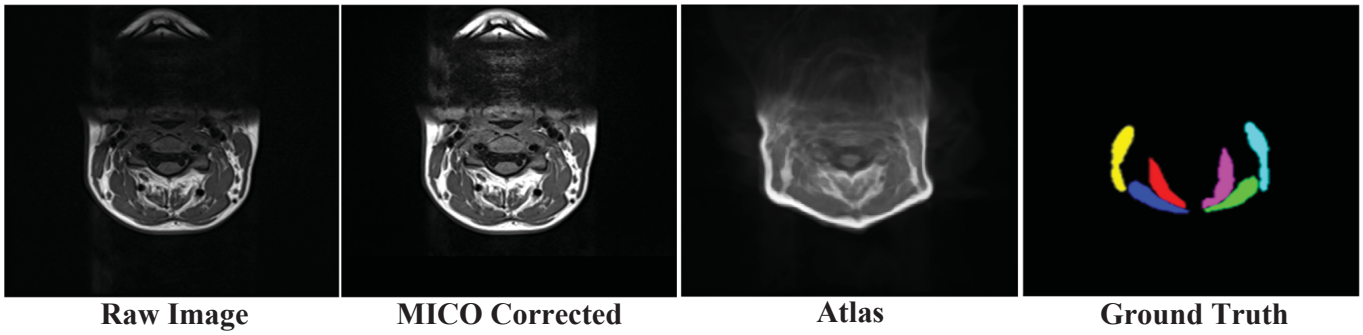

Fig. 2. Axial MR images at top of C3 spinal level showing results of inhomogeneity correction, population average and ground truth.

layer for a segmentation procedure instead of the last block which the U-Net uses. The MFP-Unet performed excellently for echocardiographic segmentation, better than the U-Net, U-Net++, deeplabv3, an anatomically constrained neural network (ACNN) and a stacked hourglass (SHG) network [20]. The R2Unet assures better feature representation than the U-Net with the same number of parameters and yielded good results compared with those of the SegNet and U-Net for segmenting retinal blood vessels, skin cancers and lung lesions [21]. The U-Net++ decreases the semantic gap between the feature maps of the encoder and decoder of the U-Net through dense nested skip connections. It provided better segmentation accuracy than the U-Net and wide U-Net for 3D CTs of chest nodules, nuclei microscopic images, liver CTs and polyp colonoscopy videos [19].

We integrated a multi-scale strategy with the U-Net in a serial and parallel way because our dataset contains neck MRIs with muscles of various sizes and shapes, and this strategy has shown good performances for anatomical variabilities [23]. It encodes both local and global contexts, textures and shapes. In the MS-Unet, we implemented this strategy in a U-Net architecture using kernels of different sizes in different blocks of contracting and expanding paths. Another popular strategy for implementing a multi-scale technique is the multi-stream architecture [23]–[25] which we used in our PMS-Unet to enable the processing of a large context through different resolutions of input images without increasing the receptive fields, thereby reducing the memory and computational requirements of deep neural networks. Three pathways were decoupled with different kernel sizes and merged in the contraction path for the first 3 convolutional blocks of the U-Net which was replicated in the expanding path for the 7th and 8th blocks in the PMS-Unet. This multi-stream concept was also extended for two different networks in the R2A-Unet++ in which the R2A-Unet and U-Net++ were considered different pathways and merged in the output layer. The attention gate was integrated with the R2Unet in the R2A-Unet to boost prediction performance.

The networks were trained by a combination of the loss functions of cross-entropy and dice coefficient using a multi-data training technique [17], in which both a gradient magnitude and training images were used, that had shown improved performances. Also, batch normalization and ReLU activation were used in each layer after convolution.

C. Implementation

The CNNs were trained on a NVIDIA GeForce GTX Titan Linux GPU (12 GB) using TensorFlow and Keras, and an Adam optimizer with a 0.001 learning rate. The images were divided into 30 training, 5 validation and 10 testing ones with almost gender equality. In addition, gradient magnitude images of the corresponding training images were used during training resulting in a total of 60 training images. Real-time data augmentation was performed to supplement the training dataset. The augmentation took the form of an affine transformation (rotation, translation, shearing and scaling) and the nearest fill mode for the image and mask. The number of epochs and batch size were set to 1500 and 5, respectively, during training. The models were initialized by drawing weights from a 0-centered truncated normal distribution.

III. Results and Analyses

Dice similarity coefficients (DSCs) and directed Hausdorff distances (DHDs) [26] were used to quantitatively analyze the segmentation results. We used the DHD because the DSC ignores the positions of the pixels which are important in segmentation. Table I shows the average DSC and DHD values achieved by the models and it can be seen that the U-Net++ yielded the best result in terms of the DSC with an overall mean of 0.89. Although the A-Unet obtained better results than the U-Net for two 3D CT abdominal datasets, it achieved a lower overall mean DSC than most of the models. On the other hand, the A-Unet and R2Unet models showed better overall mean DHDs, each with value of 1.89 mm. It is also noticeable that the PMS-Unet provided relatively good performances in terms of both metrics with its overall means of DSC and DHD 0.88 and 1.90 mm, respectively. If we consider both evaluation measures to compare performances, it can be said that the R2Unet outperformed the others. Of the muscles, M1 showed the best results from both metrics in almost all the models, possibly because of its larger size.

To further analyze the A-Unet, R2Unet, U-Net++ and PMS-Unet models, the visual segmentation results for some test patients are shown in Fig. 3. It can be seen that the R2Unet was better than the other methods and, although there were

TABLE I
AVERAGE DSC AND DHD VALUES FOR THE NETWORKS. THE BEST RESULTS IN THE OVERALL COLUMN ARE BOLDED.

	Networks	M1	M2	M3	M4	M5	M6	Overall
DSC	U-Net	0.94±0.03	0.86±0.03	0.91±0.05	0.85±0.07	0.86±0.04	0.86±0.06	0.88±0.06
	CRF-Unet	0.94±0.04	0.86±0.04	0.91±0.04	0.85±0.05	0.84±0.04	0.85±0.06	0.88±0.06
	A-Unet	0.93±0.03	0.86±0.04	0.91±0.05	0.85±0.06	0.84±0.06	0.85±0.06	0.87±0.06
	MFP-Unet	0.93±0.03	0.86±0.05	0.90±0.05	0.85±0.06	0.84±0.05	0.87±0.04	0.88±0.06
	R2Unet	0.92±0.03	0.86±0.04	0.93±0.03	0.85±0.08	0.86±0.04	0.85±0.07	0.88±0.06
	U-Net++	0.94±0.03	0.86±0.04	0.93±0.03	0.86±0.06	0.86±0.05	0.86±0.06	**0.89±0.06**
	R2A-Unet	0.94±0.03	0.85±0.06	0.92±0.04	0.84±0.07	0.86±0.06	0.85±0.06	0.88±0.07
	R2AUnet++	0.92±0.04	0.86±0.05	0.90±0.04	0.85±0.06	0.87±0.04	0.86±0.06	0.88±0.05
	PMS-Unet	0.94±0.03	0.86±0.03	0.92±0.03	0.85±0.06	0.86±0.05	0.85±0.08	0.88±0.06
	MS-Unet	0.94±0.03	0.85±0.04	0.90±0.04	0.84±0.07	0.84±0.06	0.85±0.06	0.87±0.07
DHD (mm)	U-Net	1.49±0.26	1.93±0.27	1.60±0.59	2.26±0.52	2.21±0.37	2.23±0.37	1.95±0.52
	CRF-Unet	1.49±0.26	1.96±0.26	1.72±0.36	2.36±0.46	2.26±0.28	2.33±0.26	2.02±0.46
	A-Unet	1.56±0.41	1.83±0.26	1.56±0.60	2.13±0.43	2.04±0.36	2.20±0.42	1.89±0.50
	MFP-Unet	1.51±0.15	1.83±0.26	1.64±0.55	2.36±0.50	2.38±0.39	2.18±0.38	1.98±0.52
	R2Unet	1.67±0.32	1.76±0.33	1.51±0.15	2.10±0.52	2.09±0.38	2.20±0.38	**1.89±0.45**
	U-Net++	1.46±0.27	1.89±0.38	1.54±0.47	2.25±0.57	2.08±0.27	2.37±0.40	1.93±0.53
	R2A-Unet	1.63±0.19	2.03±0.41	1.50±0.50	2.27±0.51	2.17±0.31	2.24±0.32	1.97±0.49
	R2AUnet++	1.70±0.63	2.11±0.35	1.56±0.24	2.19±0.35	2.28±0.34	2.38±0.34	2.04±0.50
	PMS-Unet	1.39±0.28	1.91±0.26	1.59±0.42	2.20±0.39	2.19±0.32	2.13±0.28	1.90±0.45
	MS-Unet	1.42±0.30	2.07±0.35	1.62±0.52	2.28±0.54	2.32±0.37	2.19±0.35	1.98±0.54

some scattered labels and holes for the M1 muscle (cyan), all the other muscles' boundaries were quite smooth except for a few holes and additional false labeling.

IV. DISCUSSION AND CONCLUSION

This study evaluated 10 (6 original and 4 modified) deep learning segmentation methods on the DSCs and DHDs of manually delineated neck muscles using different versions of the U-Net. We modified the models based on recent popular techniques for machine learning, such as multi-stream and multi-scale approaches. This type of study is unprecedented in the deep learning-based segmentation literature and, although its scale is small (the project is ongoing), we hope that it will be beneficial for the deep learning-based medical image segmentation community.

Our main finding from this study was that the performances of the models were quite similar except for the R2Unet which achieved slightly better accuracy when both metrics were combined (DSC of 0.88 and DHD of 1.89). Although the U-Net++ model achieved a better DSC of 0.89, it's DHD was 1.93. On the other hand, the DSC of R2Unet was a little smaller than that of U-Net++. Therefore, our conclusion is that R2Unet performs slightly better than U-Net++. Although some of the models reported better performances on other medical datasets in the literature, they did not perform well on our dataset; in particular, the A-Unet's performance was worse than that of the U-Net in terms of the DSC but performed well using CT data for pancreatic segmentation. The reason might be the complex anatomy associated with neck muscles while the pancreas is comparatively larger and not as closely surrounded by other tissue types. There is no recommendation to use any particular model at this stage as further research using a larger dataset needs to be conducted.

In future, we will include some additional networks with more data for other spinal levels as well as analyze more evaluation measures using a diverse dataset for 3D learning.

REFERENCES

[1] C. Lungu, A. W. Tarulli, D. Tarsy, P. Mongiovi, V. G. Vanderhorst, and S. B. Rutkove, "Quantifying muscle asymmetries in cervical dystonia with electrical impedance: a preliminary assessment," *Clinical Neurophysiology*, vol. 122, no. 5, pp. 1027–1031, 2011.

[2] J. Elliott, A. Pedler, J. Kenardy, G. Galloway, G. Jull, and M. Sterling, "The temporal development of fatty infiltrates in the neck muscles following whiplash injury: an association with pain and posttraumatic stress," *PloS one*, vol. 6, no. 6, p. e21194, 2011.

[3] S. Mesbah, A. M. Shalaby, S. Stills, A. M. Soliman, A. Willhite, S. J. Harkema, E. Rejc, and A. S. El-Baz, "Novel stochastic framework for automatic segmentation of human thigh MRI volumes and its applications in spinal cord injured individuals," *PloS one*, vol. 14, no. 5, 2019.

[4] A. Al Suman, N. Aktar, M. Asikuzzaman, A. L. Webb, D. M. Perriman, and M. R. Pickering, "Atlas-based segmentation of neck muscles from mri for the characterisation of whiplash associated disorder," in *Eighth International Conference on Digital Image Processing*, 2016, pp. 100 334L–100 334L.

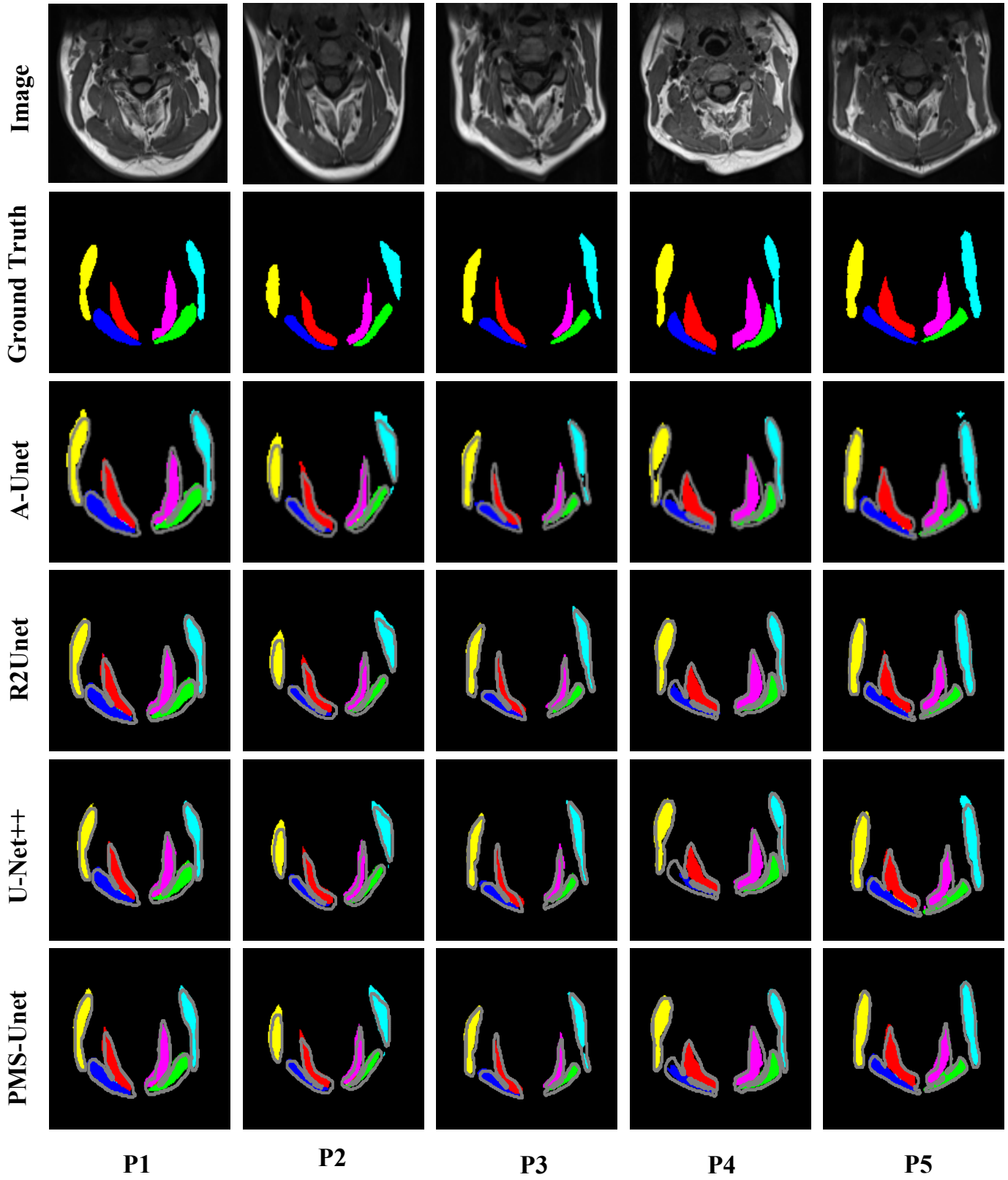

Fig. 3. Visual segmentation results for MRIs and ground truths for A-Unet, R2Unet, U-Net++ and PMS-Unet models (columns represent different patients denoted as P1, P2, P3, P4 and P5). The extracted contours from the ground truths denoted by a gray line are superimposed on the automatic segmentation results.

[5] A. A. Suman, M. N. Aktar, M. Asikuzzaman, A. L. Webb, D. M. Perriman, and M. R. Pickering, "Segmentation and reconstruction of cervical muscles using knowledge-based grouping adaptation and new step-wise registration with discrete cosines," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–13, 2017.

[6] C. Pons, B. Borotikar, M. Garetier, V. Burdin, D. B. Salem, M. Lempereur, and S. Brochard, "Quantifying skeletal muscle volume and shape in humans using MRI: A systematic review of validity and reliability," *PloS one*, vol. 13, no. 11, p. e0207847, 2018.

[7] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Drinet for medical image segmentation," *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2453–2462, 2018.

[8] 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*, 2015, pp. 234–241.

[9] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-Net: Context encoder network for 2D medical image segmentation," *IEEE transactions on medical imaging*, 2019.

[10] J. C. Caicedo, J. Roth, A. Goodman, T. Becker, K. W. Karhohs, M. Broisin, C. Molnar, C. McQuin, S. Singh, F. J. Theis *et al.*, "Evaluation of deep learning strategies for nucleus segmentation in fluorescence images," *Cytometry Part A*, vol. 95, no. 9, pp. 952–965, 2019.

[11] A. Hänsch, M. Schwier, T. Gass, T. Morgas, B. Haas, V. Dicken, H. Meine, J. Klein, and H. K. Hahn, "Evaluation of deep learning methods for parotid gland segmentation from CT images," *Journal of Medical Imaging*, vol. 6, no. 1, p. 011005, 2018.

[12] C. Li, J. C. Gore, and C. Davatzikos, "Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation," *Magnetic resonance imaging*, vol. 32, no. 7, pp. 913–923, 2014.

[13] A. Al Suman, M. Asikuzzaman, A. L. Webb, D. M. Perriman, and M. R. Pickering, "Inter-subject image registration of clinical neck MRI volumes using discrete periodic spline wavelet and free form deformation," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–5.

[14] M. Asikuzzaman, A. Al Suman, and M. R. Pickering, "Epd similarity measure and demons algorithm for object-based motion estimation," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–7.

[15] A. Al Suman, M. Asikuzzaman, A. L. Webb, D. M. Perriman, M. Tahtali, and M. R. Pickering, "A deformable 3d-3d registration framework using discrete periodic spline wavelet and edge position difference," *IEEE Access*, vol. 8, pp. 146 116–146 133, 2020.

[16] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.

[17] W. Xia, M. Fortin, J. Ahn, H. Rivaz, M. C. Battié, T. M. Peters, and Y. Xiao, "Automatic paraspinal muscle segmentation in patients with lumbar pathology using deep convolutional neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 318–325.

[18] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical image analysis*, vol. 53, pp. 197–207, 2019.

[19] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[20] S. Moradi, M. G. Oghli, A. Alizadehasl, I. Shiri, N. Oveisi, M. Oveisi, M. Maleki, and J. Dhooge, "MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography," *Physica Medica*, vol. 67, pp. 58–69, 2019.

[21] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual u-net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 1, p. 014006, 2019.

[22] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1529–1537.

[23] 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, 2017.

[24] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[25] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, "Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2421–2433, 2015.

[26] A. A. Taha and A. Hanbury, "An efficient algorithm for calculating the exact hausdorff distance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2153–2163, 2015.