Automated Contouring Using Neural Networks for 4.2.0 CT Models

Hanlin Wan, PhD
MIM Software Inc., Cleveland, OH, USA

Introduction

Automated segmentation of structures on medical images has always been challenging. In clinical practice, much of the segmentation is performed manually, which is highly time-consuming. Many semi- and fully-automated algorithms have been developed to aid physicians in contouring, such as thresholding,[1] edge-detection-based methods,[2] deformable models,[3] and atlas-based registration.[4]

However, many structures have various textural and intensity patterns, making it almost impossible for any single computer vision-based algorithm to work effectively in a universal fashion. In recent years, research on the use of neural networks has grown rapidly. The performance of these networks on many computer vision tasks has often surpassed that of a human, such as in image classification tasks like ImageNet.[5]

Contour ProtégéAI+™ provides a neural network framework for automated contouring of normal structures on CT and MR images.

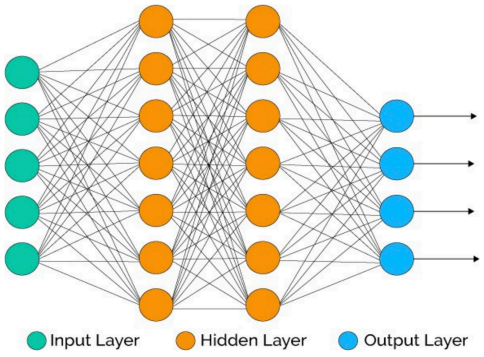

FIGURE 1. *Schematic of a basic neural network.*

Neural networks attempt to mimic how the human brain works. The brain consists of billions of neurons. Each neuron receives multiple signals from other neurons and sends out a signal based on those inputs. These neurons are often organized into layers, allowing the brain to use simple building blocks to process complicated input signals. Figure 1 shows an example of a basic neural network. Each line has a weight and an associated bias.

Therefore, each node calculates a weighted sum of its inputs and then applies some activation function. The output is then sent to the next node until the final output layer is reached. When training a neural network, pairs of inputs and desired outputs are shown. The network learns to adjust the network parameters to minimize the difference between its outputs and the desired outputs. Essentially, neural networks learn to recognize patterns much in the same way the human brain does. This pattern learning is what makes neural networks so powerful, allowing them to exceed the performance of traditional methods.

Contour ProtégéAI+'s neural network model is based on the U-Net architecture, which has been used for segmentation in numerous different applications. The model consists of many layers of weights and biases as mentioned above to transform the input image to a segmentation mask for each structure at the final output layer. This output is then post-processed to keep the single, largest connected component. Appropriate image visualization software must be used to review and, if necessary, edit results automatically generated by Contour ProtégéAI+.

Training and Validation

A large, multi-institution dataset was assembled for training, along with a separate, large multi-institutional dataset for validation. None of the validation data came from any of the institutions from the training pool. The mean and standard deviation of the model performance on this validation set was then calculated.

Results

Five different sets of metrics were used to assess the performance of the neural network segmentations:

1. Dice coefficient - a measure of the spatial overlap between the ground truth contours and the neural network segmentations.
2. Mean distance to agreement (MDA) - the mean symmetric surface distance between the ground truth contours and the neural network segmentations.
3. Qualitative user feedback score - assessment of contour quality by experienced users on a scale of 1-3 (none, moderate, significant time savings compared to contouring from scratch).
4. Localization success - percentage of images where the structure was correctly localized by the neural network segmentation.
5. Added path length (APL) - a measure of the cumulative amount of editing in mm needed to match the ground truth contour.

Tables 1-4 tabulate the Dice, MDA, user feedback score, and localization success for each of the models, and Table 5 shows the APL per model. Figures 2-9 compare the mean Dice and mean MDA of the neural network and MIM Software's atlas using majority vote 5. In all cases, Contour ProtégéAI+'s neural network segmentations were proportionate or superior to atlas-based segmentation.

TABLE 1. *Performance statistics for the CT Thorax model.*

Structure	Mean ± Std Dice	Mean ± Std MDA (mm)	Beta User Feedback Score	Localization Success on Relevant FOV CTs	Localization Success on Whole Body CTs
BrachialPlex_L	0.41 ± 0.15	2.89 ± 1.07	2.43	100	100
BrachialPlex_R	0.39 ± 0.15	3.11 ± 1.38	2.43	100	100
Breast_L	0.79 ± 0.07	5.00 ± 2.53	2.38	100	100
Breast_R	0.80 ± 0.11	4.49 ± 2.02	2.57	100	100
Breast_L_RTOG	0.77 ± 0.11	5.47 ± 3.72	2.50	100	100
Breast_R_RTOG	0.80 ± 0.15	5.10 ± 3.17	2.50	100	100
Bronchus	0.62 ± 0.14	1.86 ± 0.90	2.63	100	100
Carina	0.50 ± 0.12	1.93 ± 0.83	2.43	99	100
Cricoid	0.06 ± 0.05	5.01 ± 1.40	2.86	91	100
Esophagus	0.69 ± 0.16	1.08 ± 1.53	2.50	99	100
Glnd_Thyroid	0.66 ± 0.18	1.68 ± 1.44	2.86	100	77
GreatVes	0.70 ± 0.16	3.44 ± 1.99	2.63	100	91
Heart	0.90 ± 0.07	2.49 ± 1.84	2.63	100	100
Humerus_Head_L	0.62 ± 0.24	0.33 ± 0.22	3.00	100	100
Humerus_Head_R	0.60 ± 0.22	0.33 ± 0.35	3.00	100	100
Kidney_L	0.90 ± 0.09	1.32 ± 1.13	2.75	100	95
Kidney_R	0.89 ± 0.09	1.41 ± 1.06	2.75	100	100
Larynx	0.59 ± 0.14	2.87 ± 1.34	2.63	100	100
Liver	0.90 ± 0.13	3.56 ± 9.12	2.71	99	95
Lung_L	0.96 ± 0.02	0.79 ± 0.40	2.75	100	100
Lung_R	0.97 ± 0.03	0.85 ± 0.48	2.75	100	100
Musc_Constrict	0.50 ± 0.17	1.67 ± 1.55	3.00	100	91
Pancreas	0.47 ± 0.21	6.80 ± 8.89	2.17	96	95
SpinalCord	0.64 ± 0.17	1.28 ± 0.72	2.50	100	100
Stomach	0.73 ± 0.21	6.89 ± 20.57	2.13	97	100
Trachea	0.74 ± 0.17	1.12 ± 0.64	2.63	99	100
A_LAD	0.32 ± 0.12	3.43 ± 4.46	2.14	100	86
A_Aorta_Asc	0.83 ± 0.17	1.11 ± 0.60	2.67	96	100
Rib	0.28 ± 0.11	39.04 ± 14.60	2.63	100	86
Chestwall_L	0.39 ± 0.17	4.44 ± 1.24	2.43	100	100
Chestwall_R	0.42 ± 0.18	4.69 ± 1.56	2.43	100	100
LN_Ax_L1_L	0.63 ± 0.10	2.25 ± 0.80	2.67	100	100
LN_Ax_L1_R	0.58 ± 0.13	2.90 ± 1.52	2.67	100	100
LN_Ax_L2_L	0.64 ± 0.13	1.93 ± 0.88	2.67	100	100
LN_Ax_L2_R	0.60 ± 0.18	2.44 ± 1.26	2.67	100	100
LN_Ax_L3_L	0.62 ± 0.19	1.90 ± 1.26	2.67	100	100

Continued on next page

TABLE 1 CONT. *Performance statistics for the CT Thorax model.*

Structure	Mean ± Std Dice	Mean ± Std MDA (mm)	Beta User Feedback Score	Localization Success on Relevant FOV CTs	Localization Success on Whole Body CTs
LN_Ax_L3_R	0.52 ± 0.17	2.74 ± 1.73	2.67	100	100
LN_IMN_L	0.41 ± 0.17	1.82 ± 0.85	3.00	100	100
LN_IMN_R	0.48 ± 0.20	1.69 ± 1.38	3.00	100	100
LN_Sclav_L	0.66 ± 0.13	2.49 ± 1.70	2.00	100	100
LN_Sclav_R	0.55 ± 0.09	2.67 ± 0.90	2.33	100	100

FIGURE 2. *Comparison of atlas and Contour ProtégéAI+ mean Dice for the CT Thorax model.*

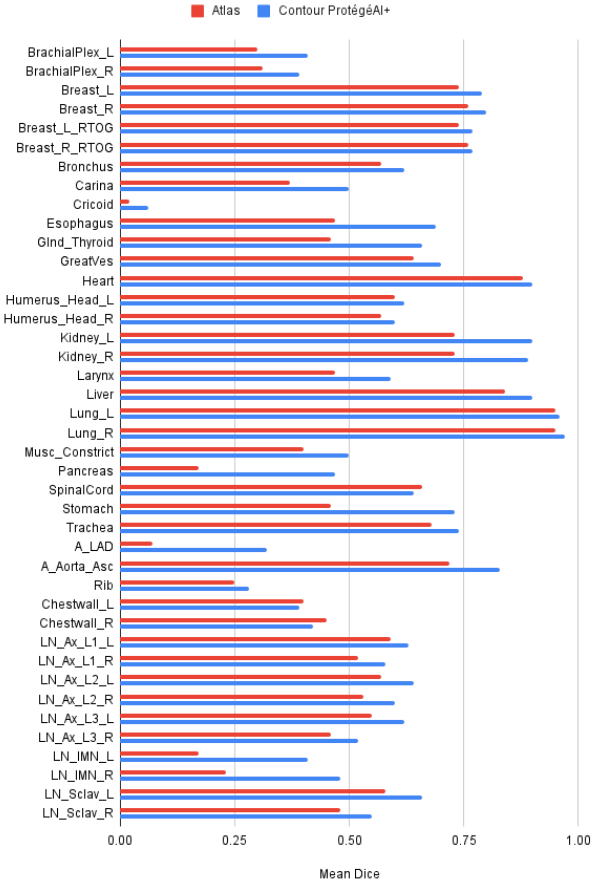

FIGURE 3. *Comparison of atlas and Contour ProtégéAI+ mean MDA for the CT Thorax model.*

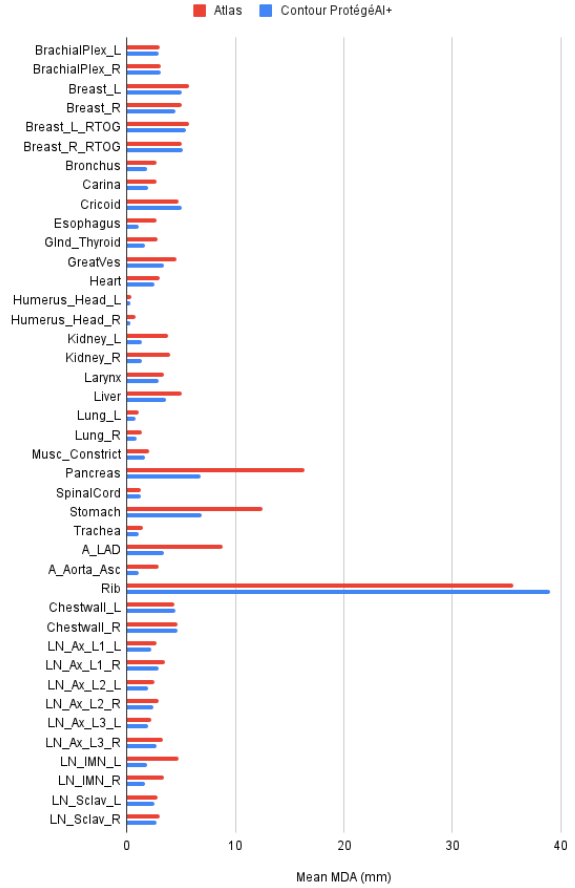

TABLE 2. *Performance statistics for the CT Abdomen model.*

Structure	Mean ± Std Dice	Mean ± Std MDA (mm)	Beta User Feedback Score	Localization Success on Relevant FOV CTs	Localization Success on Whole Body CTs
Bladder	0.92 ± 0.16	0.78 ± 0.73	2.60	98	95
Bowel	0.52 ± 0.19	5.05 ± 3.17	2.50	100	100
BowelBag	0.36 ± 0.11	10.08 ± 3.36	2.89	100	100
CaudaEquina	0.69 ± 0.13	0.95 ± 0.53	2.60	100	100
Kidney_L	0.94 ± 0.03	0.82 ± 0.41	2.90	100	91
Kidney_R	0.92 ± 0.07	0.96 ± 0.62	2.90	100	100
Liver	0.93 ± 0.08	2.01 ± 2.14	2.80	100	100
SpinalCord	0.65 ± 0.14	0.83 ± 0.32	2.90	100	100
Stomach	0.82 ± 0.11	2.68 ± 2.02	2.50	100	100

FIGURE 4. *Comparison of atlas and Contour ProtégéAI+ mean Dice for the CT Abdomen model.*

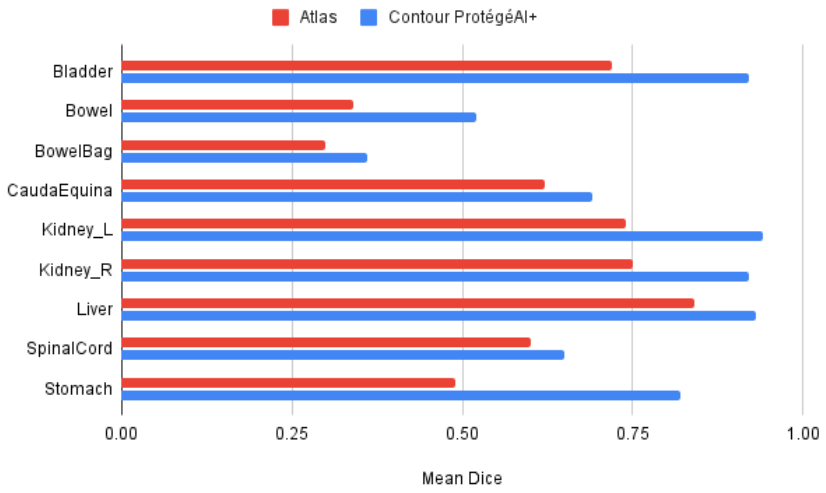

FIGURE 5. *Comparison of atlas and Contour ProtégéAI+ mean MDA for the CT Abdomen model.*

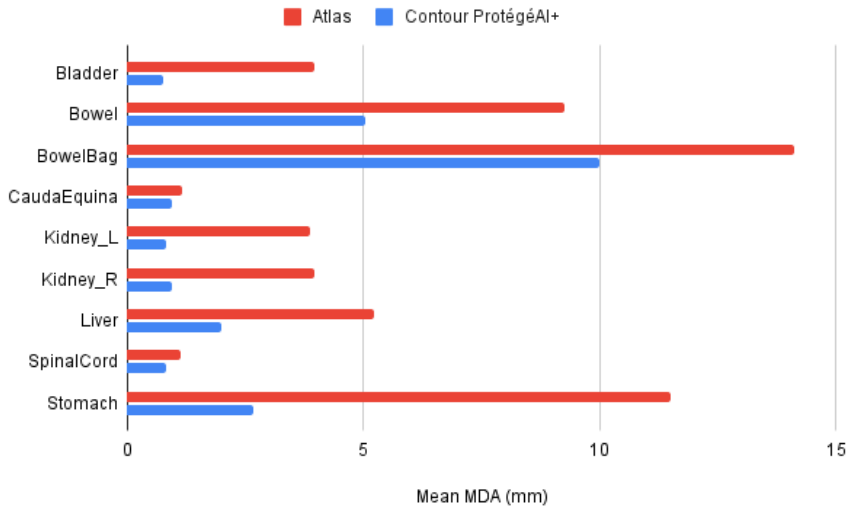

TABLE 3. *Performance statistics for the CT Female Pelvis model.*

Structure	Mean ± Std Dice	Mean ± Std MDA (mm)	Beta User Feedback Score	Localization Success on Relevant FOV CTs	Localization Success on Whole Body CTs
Bladder	0.91 ± 0.06	1.04 ± 0.94	2.75	100	100
Bag_Bowel	0.52 ± 0.20	6.89 ± 2.25	2.63	100	100
Bowel	0.55 ± 0.19	4.84 ± 3.03	2.75	100	100
Colon_Sigmoid	0.47 ± 0.24	14.72 ± 15.10	2.50	93	86
Femur_Head_L	0.91 ± 0.08	0.95 ± 1.03	2.75	100	100
Femur_Head_R	0.92 ± 0.02	0.69 ± 0.28	2.75	100	95
UteroCervix	0.65 ± 0.27	4.39 ± 13.21	2.00	97	100
LN_Pelvics	0.74 ± 0.06	4.12 ± 1.52	2.75	100	100
Rectum	0.75 ± 0.12	1.52 ± 1.66	2.63	100	100
SacralPlex	0.03 ± 0.01	12.81 ± 1.86	2.50	100	100
Sacrum	0.89 ± 0.01	1.09 ± 0.17	3.00	100	100
CaudaEquina	0.66 ± 0.11	0.94 ± 0.59	2.57	100	100

FIGURE 6. *Comparison of atlas and Contour ProtégéAI+ mean Dice for the CT Female Pelvis model.*

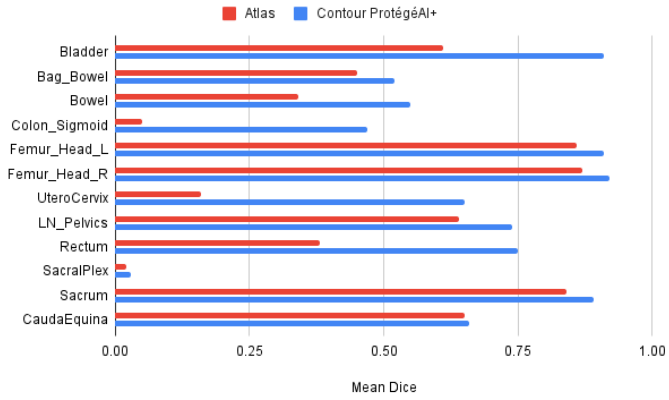

FIGURE 7. *Comparison of atlas and Contour ProtégéAI+ mean MDA for the CT Female Pelvis model.*

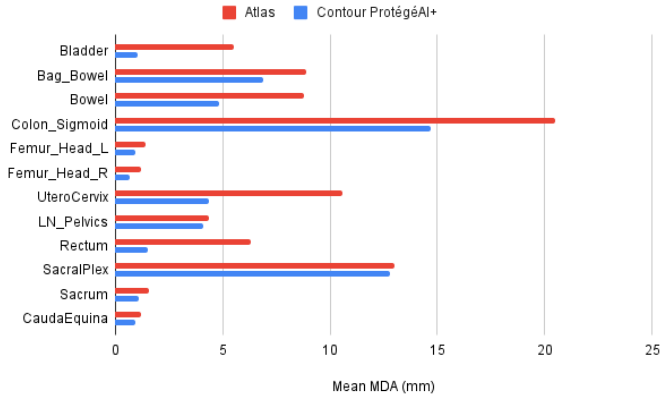

TABLE 4. *Performance statistics for the CT SurePlan MRT model.*

Structure	Mean ± Std Dice	Mean ± Std MDA (mm)	Beta User Feedback Score	Localization Success on Relevant FOV CTs	Localization Success on Whole Body CTs
Bone	0.83 ± 0.05	4.57 ± 3.35	2.67	100	100
Glnd_Lacrimal_L	0.30 ± 0.21	1.26 ± 0.57	2.67	73	100
Glnd_Lacrimal_R	0.36 ± 0.23	1.18 ± 0.87	2.67	76	95
Glnd_Submand_L	0.67 ± 0.29	1.00 ± 0.35	3.00	85	95
Glnd_Submand_R	0.66 ± 0.31	0.97 ± 0.33	3.00	83	100
Glnd_Thyroid	0.75 ± 0.12	1.29 ± 1.20	3.00	100	82
Kidney_L	0.91 ± 0.04	1.56 ± 0.61	3.00	100	91
Kidney_R	0.91 ± 0.03	1.45 ± 0.63	3.00	100	95
Liver	0.93 ± 0.07	1.79 ± 1.50	3.00	100	100
Lung_L	0.96 ± 0.04	0.90 ± 0.49	3.00	100	100
Lung_R	0.96 ± 0.05	1.07 ± 0.86	3.00	100	100
Parotid_L	0.81 ± 0.05	1.42 ± 0.40	3.00	100	100
Parotid_R	0.82 ± 0.05	1.34 ± 0.48	3.00	100	100
Spleen	0.95 ± 0.02	0.62 ± 0.45	3.00	100	100

FIGURE 8. *Comparison of atlas and Contour ProtégéAI+ mean Dice for the CT SurePlan MRT model.*

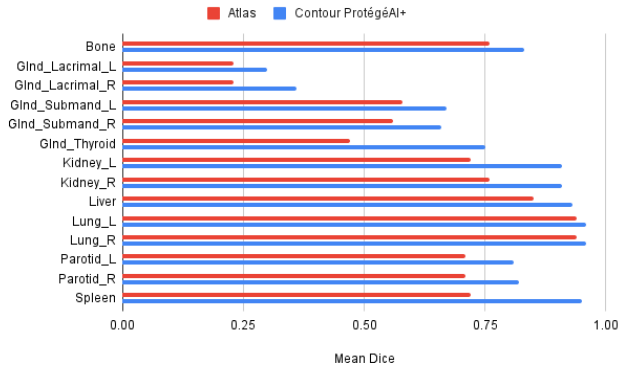

FIGURE 9. *Comparison of atlas and Contour ProtégéAI+ mean MDA for the CT SurePlan MRT model.*

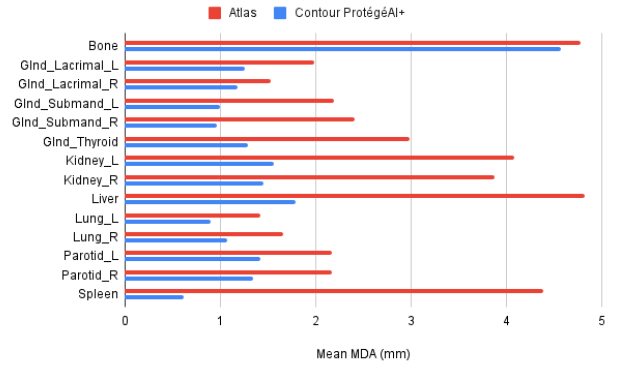

TABLE 5. *Comparison of atlas and Contour ProtégéAI+ mean added path length (mm).*

Model	Atlas	Contour ProtégéAI+
Thorax	220.33 ± 232.42	181.44 ± 219.43
Abdomen	433.24 ± 392.19	314.13 ± 349.77
Female Pelvis	354.53 ± 386.05	238.05 ± 309.40
SurePlan MRT	216.52 ± 207.87	133.01 ± 160.23

References

1. Weszka JS. A survey of threshold selection techniques. *Comput Graph Image Process*. 1978;7(2):259-265. doi:10.1016/0146-664X(78)90116-8
2. Senthilkumaran N, Rajesh R. Edge detection techniques for image segmentation-a survey of soft computing approaches. *Int J Recent Trends Eng*. 2009;1(2):250.
3. McInerney T, Terzopoulos D. Deformable models in medical image analysis: a survey. *Med Img Anal*. 1996;1(2):91-108. doi: 10.1016/S1361-8415(96)80007-7
4. Klein S, Staring M, Murphy K, et al. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2009;29(1):196-205. doi: 10.1109/TMI.2009.2035616
5. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211-252. doi: 10.1007/s11263-015-0816-y