

# MISS-tool: Medical Image Segmentation Synthesis Tool to Emulate Segmentation Errors

Shuyue Guan, Ravi K. Samala, Arian Arab, and Weijie Chen

Division of Imaging, Diagnostics, and Software Reliability  
Office of Science and Engineering Laboratories  
Center for Devices and Radiologic Health

United States Food and Drug Administration, Silver Spring, Maryland, United States  
{shuyue.guan, ravi.samala, arian.arab, weijie.chen}@fda.hhs.gov

## ABSTRACT

Segmentation of medical images with known ground truth is useful for investigating properties of performance metrics and comparing different approaches of combining multiple manual segmentations to establish a reference standard, thereby informing selection of performance metrics and truthing methods. For medical images, however, segmentation ground truth is typically not available. One way of synthesizing segmentation errors is to use regular geometric objects as ground truth, but they lack the complexity and variability of real anatomical objects. To address this problem, we developed a medical image segmentation synthesis (MISS)-tool. The MISS-tool emulates segmentations by adjusting truth masks of anatomical objects extracted from real medical images. We categorized six types of segmentation errors and developed contour transformation tools with a set of user-adjustable parameters to modify the defined truth contours to emulate different types of segmentation errors, thereby generating synthetic segmentations. In a simulation study, we synthesized multiple segmentations to emulate algorithms or observers with pre-defined sets of segmentation errors (*e.g.*, under/over-segmentation) using 220 lung nodule cases from the LIDC lung computed tomography dataset. We verified that the synthetic segmentation results manifest the type of errors that are consistent with our pre-configured setting. Our tool is useful for synthesizing a range of segmentation errors within a clinical segmentation task.

**Keywords:** Medical Image Segmentation Synthesis, Manual Segmentation Emulation, Segmentation Evaluation, Segmentation Errors

## 1. INTRODUCTION

Medical image segmentation is widely used to identify the extent of anatomical objects in medical images (*e.g.*, organs or lesions) for a variety of clinical applications, such as feature extraction for diagnosis, radiation treatment planning, monitoring tumor size in response to treatment, and so on. Although cutting-edge artificial intelligence/machine learning (AI/ML) algorithms are being developed at an ever-increasing pace for medical image segmentation,<sup>1</sup> there is a lack of consensus and standardized assessment methodologies for image segmentation. For example, there are many segmentation performance metrics, but there is no consensus on methods for choosing performance metrics that are most suitable for a specific clinical task.<sup>2,3</sup> Moreover, there is no consensus in the approach of establishing a reference standard based on manual segmentation.<sup>4,5</sup>

Segmentation of medical images with known ground truth is useful to investigate properties of performance metrics, thereby informing metric selection, and to investigate different approaches of combining multiple manual segmentations to establishing a reference standard, thereby informing truthing method selection. However, identifying the true extent of an object’s boundary in real images is difficult. That causes the establishment of the ground truth for segmentation tasks a challenging process.. The use of regular geometric shapes can be useful in illustrating certain properties of performance metrics;<sup>3,6</sup> however, they lack the complexity and variability of biological characteristics (*e.g.*, organs and lesions). Using manual segmentations or manually manipulated segmentations is usually difficult due to the high cost of acquisition or the limited number of segmentations. An alternate solution, is to exploit deep learning models to generate various segmentation errors, however, this could be difficult in systematically controlling the type and extend of segmentation errors needed for a specific clinical

task. The purpose of this work is to develop a tool capable of creating synthetic segmentations based on organ or lesion contours extracted from real medical images, which we call the *Medical Image Segmentation Synthesis tool* (MISS-tool). The MISS-tool allows users to customize segmentation errors by adjustable parameters.

## 2. METHODS

According to the study by Taha et al.<sup>3</sup> and our observation of the LIDC-IDRI dataset,<sup>7</sup> we summarized six types of segmentation errors. We further developed methods to synthesize these errors based on pre-defined truth contours. Figure 1 shows illustrations of these segmentation errors.

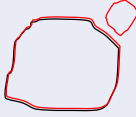
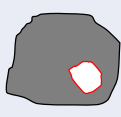
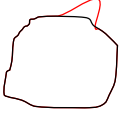
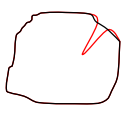
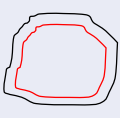
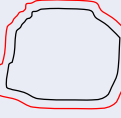
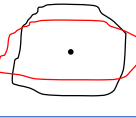
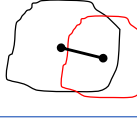
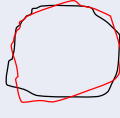
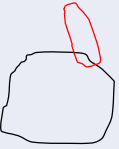

	Types of errors	Emulation Method	Illustration	
1	Satellite Structure (inside/outside)	Copy and paste with affine transformation		
2	Boundary Spiculation (outward/inward)	Contour adjustment in polar coordinate system		
3	Under/Over-size	Affine transformation		
4	Location Error (same/different)	Affine transformation		
5	Lack of Overlap (small/large)	Affine transformation		
6	Shape & Alignment	Fourier Descriptor (FD)		

Figure 1. Overview of segmentation errors and their emulation methods. Contours in black color are the truth segmentation, and contours in red color are synthetic segmentations with various errors. (1) satellite structure; the image contains small disconnected components outside/inside the main segmentation. (2) boundary spiculation error; the spike-like shape is added or removed from the segmentation. (3) under- or over-size error. (4) location error. (5) lack of overlap. (6) boundary detail error.

To generate synthetic segmentations using our MISS-tool, one must first define truth masks or contours of the objects of interest, *e.g.*, lesions, organs at risk, *etc.*. This can be achieved by manual segmentation of those objects from medical images by an expert. If multiple manual segmentations for each object are available (*e.g.*, by a panel of experts), they can be combined into one truth mask using some truthing methods such as the *Simultaneous Truth And Performance Level Estimation* (STAPLE).<sup>4</sup> Note that the "truth" mask here is only for the purpose of segmentation synthesis serving as reference for the synthesized segmentation errors. Then for each object, one can emulate segmentations with different types of segmentation errors as shown in Figure 1, using a combination of four methods (Figure 2):

- **Affine transformation (Sec. 2.1):** resizing, shifting, rotate, *etc.*
- **Add satellite structures (Sec. 2.2):** relatively small areas disconnected from the main segmentation
- **Add spiculation (Sec. 2.3):** adjust contours in the polar coordinate system
- **Fourier Descriptor (Sec. 2.4):** add, remove, or change components in the spatial frequency domain

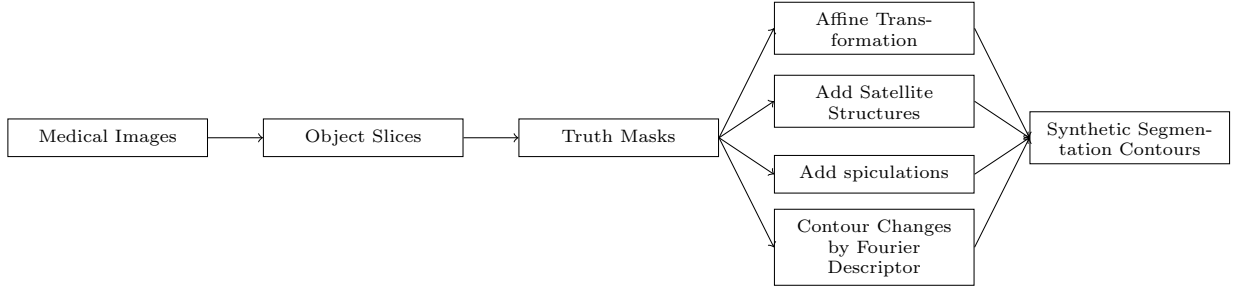


Figure 2. Flowchart of synthetic medical image segmentation generation

## 2.1 Affine Transformation

The segmentation contours can be modified by an affine transformation. As shown in Figure 3, the MISS-tool provides three types of affine transformations: resizing (changing the height and width ratio), shifting, and rotation. Resizing has two parameters: 1) the ratio of height  $R_y$  and 2) the ratio of width  $R_x$ ; a ratio equal to 1 means no change. Shifting has two parameters for 2-D images characterizing the location change in  $(x, y)$  coordinates. The rotation has one angle parameter  $\varphi$ .

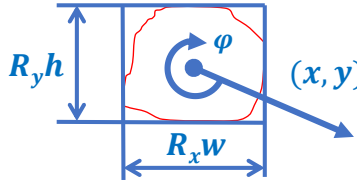


Figure 3. Affine transformation to modify the segmentation contour (in red).  $h$  and  $w$  are the height and width of the bounding box of the contour.

## 2.2 Satellite Structure Synthesis

Satellite structure is a type of segmentation error that a separate object (typically small) in the nearby region of the true object is segmented as part of the segmentation result.

The approach we used, as shown in Figure 1(1), is to extract satellite structures from the real dataset (e.g., the LIDC-IDRI dataset as our use-case application) to create a library of satellite structures based on the manual segmentation reference. To emulate the satellite structure error, we randomly select an object from the library dataset and add it at a location nearby the truth mask with or without some affine transformation (Figure 4 right). As shown in Figure 4 left, the location of the satellite structure is determined by four parameters: a range of distance  $(a, b)$  between the center of the main segmentation and the center of the satellite and a range of angles  $(r, s)$ . By definition,  $a$  and  $b$  must be greater than the sum radius of the main segmentation and the satellite structure to avoid overlapping. Finally, the satellite structure is added to a location randomly chosen in the user-specified region.

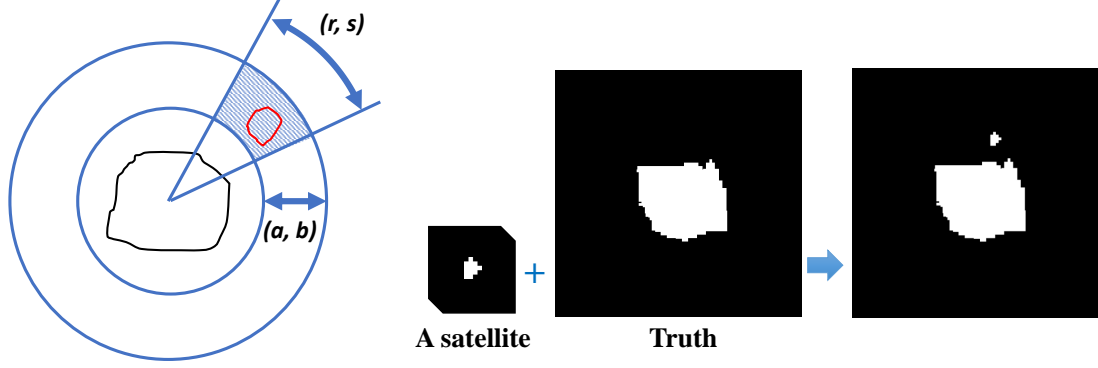


Figure 4. Left: a satellite structure (red) is randomly put aside the main segmentation (black) on the locations (dot area) defined by a range of distances  $(a, b)$  and a range of angles  $(r, s)$  from the main segmentation. Right: a real example of synthesizing satellite structures for Figure 1(1): to add the satellite structure on a truth segmentation image (positive condition).

### 2.3 Spiculations

In the polar coordinate system, the segmentation contour can be represented by the distances  $(\rho)$  and angles  $(\varphi)$  between its center of mass and points on the contour. As shown in Figure 5, point  $i$  on the contour is  $(\rho_i, \varphi_i)$  in polar coordinates. The contour can be represented by the center-to-contour distance as a function of the angle:  $\rho(\varphi)$ . The contour can be modified by changing the value of distance in a polar coordinate system. The MISS-tool uses a Gaussian function (Equation 1) to modify the truth contour to emulate boundary spiculation errors:

$$G(\varphi) = he^{-\left(\frac{\varphi-c}{w}\right)^2} \quad (1)$$

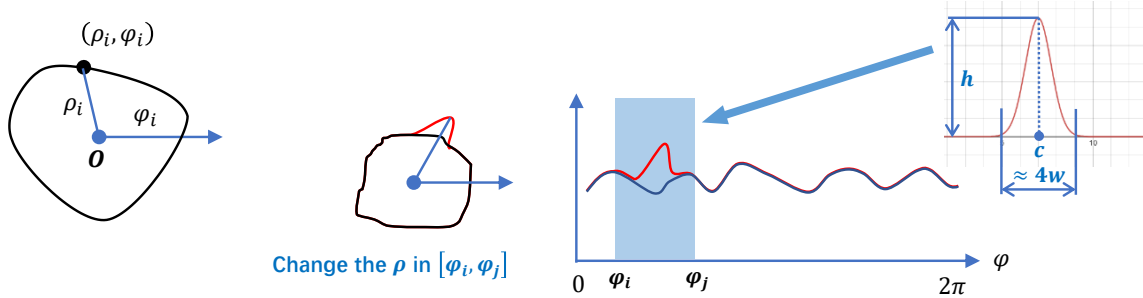


Figure 5. Emulation of spiculation using the polar coordinate system representation of the segmentation contour.

This emulation has three tunable parameters to set the center  $(c)$ , height  $(h)$ , and width  $(w)$  of each spiculation. The height can be a positive or negative value to synthesize convex or concave spiculation, as shown in Figure 1(2). The center and width are angle degrees ranging in  $[0, 2\pi]$ , and as shown in Figure 5, the real width of the spiculation (Gaussian function) is about four times the width value  $(w)$ . The contour after adding a spiculation is:

$$\rho'(\varphi) = \rho(\varphi) + G(\varphi)$$

Figure 6 shows four examples of emulated spiculation errors. Multiple instances of spiculation error can be emulated for a truth contour.



Figure 6. Examples of spiculation error, where emulated boundary spiculation are highlighted by red circles. The three parameters under each emulation from left to right are the center ( $c$ ), height ( $h$ ), and width ( $w$ ). The unit is the pixel. The height can be a positive or negative value to synthesize convex or concave spiculation.

## 2.4 Fourier Descriptor

Similar to adding spiculations, each pixel point  $k$  on the contour  $(x_k, y_k)$  in the Cartesian coordinate system can be written as a complex number:

$$p(k) = x_k + y_k i,$$

where  $k = 0, 1, 2, \dots, N - 1$ , and  $N$  is the number of points (pixels) on the contour. By using Discrete Fourier Transform (DFT):

$$f(u) = \frac{1}{N} \sum_{k=0}^{N-1} p(k) e^{-i \frac{2\pi}{N} uk},$$

where  $f(u)$  is a Fourier descriptor (FD) of the contour, and  $u = 0, 1, 2, \dots, N - 1$ . The complex number of real point  $p(k)$  can be retrieved from the FD by applying inverse-DFT:

$$p(k) \xrightleftharpoons[\text{i-DFT}]{\text{DFT}} f(u)$$

The contour can be modified if we change FDs before applying inverse-DFT. FDs describe the contour in the spatial frequency domain, ranging from low frequency (including the DC component) to high frequency. In emulating segmentation errors, the MISS-tool keeps low-frequency FDs (including the DC component) of the truth contour and allows the user to change middle-frequency FDs and remove some high-frequency FDs. This assumes a reasonable segmentation would keep the basic shape of the truth contour but may have errors in some details. This is achieved by tuning three parameters regarding thresholds of frequencies and the magnitude of changes.

Figure 7 shows an example of Fourier descriptors from a contour. Each descriptor  $f(u)$  is a complex number:

$$f(u) = \alpha_u + \beta_u i. \quad (2)$$

The plot in Figure 7 displays their magnitude:

$$|f(u)| = |\alpha_u + \beta_u i| = \sqrt{\alpha_u^2 + \beta_u^2}$$

FDs describe the components of contour in the frequency domain, and the contour can be modified by change FDs before applying inverse-DFT. Equation 2 shows an FD  $f(u)$  has two coefficients: the real part  $\alpha_u$  and the imaginary part  $\beta_u$ . Changing the values of  $\alpha_u$  and  $\beta_u$  will affect the FD and its magnitude  $|f(u)|$ .

In our MISS tool, there are three adjustable parameters that can be tuned to generate a segmentation contour from the truth contour. As shown in Figure 7, the first parameter is **Detail**, which is the number of non-zero FDs, and the high-frequency FDs beyond those are removed (red areas in Figure 7). The second parameter is **Range**, which determines the range of FDs for modifications (blue areas in Figure 7). The bandwidth with frequencies smaller than **Range** is not changed (black area in Figure 7). The third parameter is **Magnitude** for

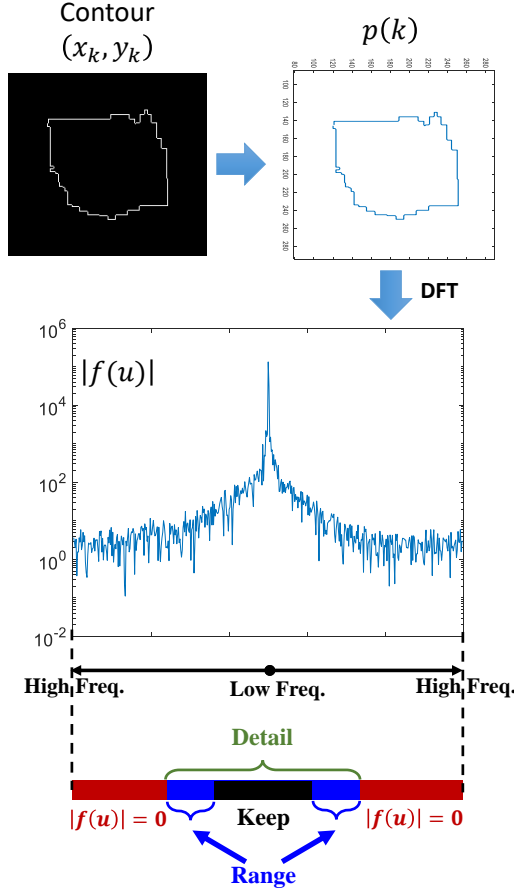


Figure 7. An example of converting the contour to Fourier descriptors. The plot displays the magnitude of Fourier descriptors re-ordered by frequency. The *Direct Current* (DC)-term:  $f(0)$  is in the center; from center to the left and right are low-frequency components to high-frequency components, which are  $f(n-1), f(N-2), \dots, f(N/2)$  and  $f(1), f(2), \dots, f(N/2)$ .

controlling how the FDs in the middle range are modified. Specifically, the real part  $\alpha_u$  and imaginary part  $\beta_u$  are changed by:

$$\begin{cases} \alpha'_u = \alpha_u + r_u m \\ \beta'_u = \beta_u + s_u m \end{cases},$$

where updated  $f'(u) = \alpha'_u + \beta'_u i$ ;  $r_u, s_u \in (-0.5, 0.5)$  are two random variables;  $m$  is the **Magnitude** parameter. Obviously,  $m = 0$  means no change to the **Range** FDs. Figure 8 shows four examples of shape and alignment changes to segmentations by processing FD.

### 3. RESULTS

We developed the MISS-tool in MATLAB to generate synthetic segmentations by modifying pre-defined truth contours as explained in the methods section. Examples are shown in Figure 9. To synthesize a segmentation contour from a pre-defined truth contour, one can use a single method or a combination of these methods (one method can be repeatedly applied with different parameters). The MISS-tool defines the basic parameters to modify the segmentation for each method, and the combination of methods provides a great amount of flexibility and variability for medical image segmentation synthesis. We also provided a graphical user interface (GUI) for tuning the parameters of these methods, as shown in Figure 10.

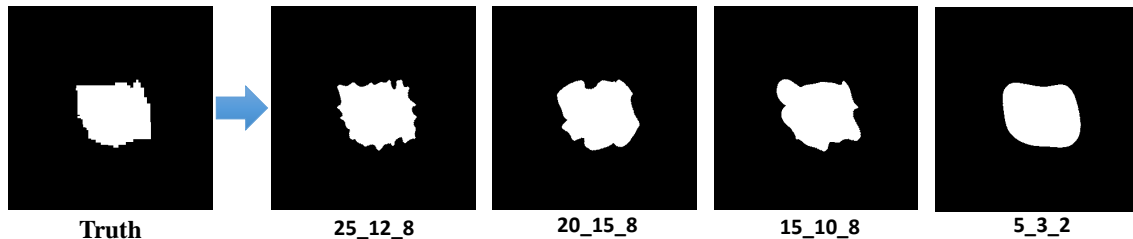


Figure 8. Examples of segmentation changes by processing FD. The three parameters from left to right are **Detail**, **Range**, and **Magnitude**, which are defined in the Method section.



Figure 9. First figure from the left is the truth mask of a lung nodule from the LIDC-IDRI dataset; others are results of MISS-tool – from left to right: results of affine transformation, satellite structure synthesis, adding boundary spiculation, Fourier descriptor, and combinations of some of the four methods. The right hand-side figures shown with the blue color highlight the truth segmentation and the synthetic segmentation shown with green contours.

To verify that the MISS-tool can generate synthetic segmentations with pre-specified segmentation errors, we synthesized multiple segmentations with pre-defined sets of segmentation errors (e.g., under/over-segmentation) on 220 independent lung nodule cases from the LIDC-IDRI dataset.<sup>7,8</sup> We used the central slice image from each nodule based on radiologist annotations. To define a truth mask, segmentation of lung nodules by multiple radiologists were combined using the STAPLE method.<sup>4,5</sup> We applied affine transformation (Sec. 2.1), which involved parameters for resizing and shifting. We tuned these parameters on 20 cases (development set) towards the pre-set segmentation errors and applied the transformations to 200 independent cases (independent test set). The goal here is to emulate over or under-segmentation (size change from -30% to +30%) by setting resizing parameters (the ratio of height  $R_y$  and width  $R_x$ ) and shifting their locations (0 or 10 pixels) using the shifting parameters  $(x, y)$ .

We quantified the segmentation accuracy of the synthetic segmentations using the size-change ratio and Intersection over Union (IoU). The size-change ratio is:

$$\text{Size-change ratio} = \frac{S_{\text{change}}}{S_{\text{truth}}},$$

where  $S_{\text{truth}}$  is the number of total pixels of the pre-defined truth mask, and  $S_{\text{seg}}$  is the number of total pixels of the synthetic segmentation. The IoU (Jaccard index)<sup>9</sup> of two segmentation masks  $A$  and  $B$  is:

$$\text{IoU}(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

To examine the potential training effect, we computed these metrics on both the development dataset and the independent test dataset.

Table 1 shows results from this preliminary verification experiment. The results verified that the synthetic segmentation results have the types and amount of errors that are consistent with the pre-set goals. The training effect of tuning the parameters appears to be small. This experiment is preliminary as it only tested tuning a subset of the tunable parameters available in the MISS-tool. Additional experiments are needed to further verify the full MISS tool.

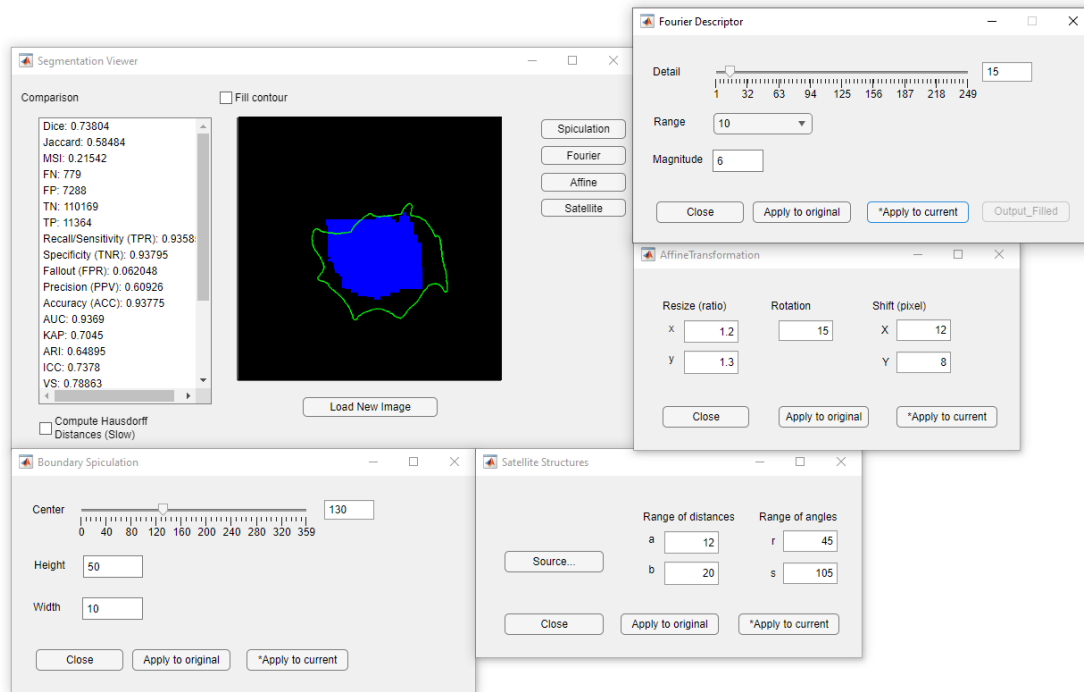


Figure 10. GUI of the MISS-tool for generating synthetic segmentations. The main window (upper-left) provides the segmentation evaluation results of 24 metric to quantify the accuracy of synthetic segmentation. The truth mask is shown in the blue area, and the green contour shows the synthesized segmentation.

Table 1. Verification of achieving pre-specified goals on segmentation errors by tuning MISS-tool parameters. For resizing  $R_y = R_x$ , and for shifting  $x = y$ . Dev 20 and Val 200: results from the development set of 20 cases and validation set of 200 cases. The results of the size-change ratio and IoU are averaged through cases in the development set and validation set respectively. Abs diff: the absolute difference between the two metric values computed on the development and validation datasets.

Goals	Changes		mean Size-change ratio			mean IoU		
	Resize	Shift	Dev 20	Val 200	Abs diff	Dev 20	Val 200	Abs diff
+30%	1.16	0	1.299	1.293	0.006	0.764	0.766	0.003
+10%	1.07	0	1.105	1.101	0.004	0.894	0.894	0.000
-10%	0.97	0	0.898	0.897	0.000	0.896	0.896	0.001
-30%	0.86	0	0.703	0.695	0.007	0.700	0.694	0.006
+30%	1.16	10	1.299	1.293	0.006	0.606	0.594	0.012
+10%	1.07	10	1.105	1.101	0.004	0.637	0.617	0.019
-10%	0.97	10	0.898	0.897	0.000	0.618	0.598	0.020
-30%	0.86	10	0.703	0.695	0.008	0.566	0.546	0.020

#### 4. CONCLUSION

We developed the Medical Image Segmentation Synthesis (MISS) tool to generate synthetic segmentations with various segmentation errors based on pre-defined ground truth masks extracted from real images. The MISS-tool emulates a variety of segmentation errors with a set of user-adjustable parameters. In a preliminary verification study, we synthesized multiple segmentations to emulate algorithms or observers with a pre-defined set of segmentation errors on lung nodule cases from the LIDC-IDRI dataset. This tool is useful for synthesizing segmentations with a range of segmentation errors that are realistic within a clinical task.

The MISS-tool can potentially be used in several ways in the development of assessment methodologies for



medical image segmentation. Synthetic segmentation can be used to investigate the properties of performance metrics, thereby informing the choice of metrics. The use of regular-shape objects, as found in the current literature, lacks the variability and complexity of the real anatomies and segregation errors found in actual medical images. The MISS-tool can also be used to simulate manual segmentations by multiple observers, which can be used to investigate methods for defining a reference standard from manual segmentations. Moreover, it can be used to investigate the impact of segmentation errors on radiomic features of lesions or dosimetric of organs at risk in a simulated radiation treatment plan.

## ACKNOWLEDGMENTS

This work was supported in part by the FDA Office of Chief Scientist and in part by an appointment of Shuyue Guan to the ORISE Research Participation Program at the Center for Devices and Radiological Health, U.S. Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and FDA/CDRH.

## REFERENCES

- [1] Hesamian, M. H., Jia, W., He, X., and Kennedy, P., “Deep learning techniques for medical image segmentation: achievements and challenges,” *Journal of digital imaging* **32**(4), 582–596 (2019).
- [2] Taha, A. A., Hanbury, A., and Toro, O. A. J. d., “A formal method for selecting evaluation metrics for image segmentation,” in [2014 *IEEE International Conference on Image Processing (ICIP)*], 932–936.
- [3] Taha, A. A. and Hanbury, A., “Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool,” *BMC medical imaging* **15**(1), 1–28 (2015).
- [4] Warfield, S. K., Zou, K. H., and Wells, W. M., “Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation,” *IEEE transactions on medical imaging* **23**(7), 903–921 (2004).
- [5] Zheng, Y., Li, G., Li, Y., Shan, C., and Cheng, R., “Truth inference in crowdsourcing: Is the problem solved?,” *Proceedings of the VLDB Endowment* **10**(5), 541–552 (2017).
- [6] Kim, H., Monroe, J. I., Lo, S., Yao, M., Harari, P. M., Machtay, M., and Sohn, J. W., “Quantitative evaluation of image segmentation incorporating medical consideration functions,” *Medical physics* **42**(6Part1), 3013–3023 (2015).
- [7] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Clarke, L. P., et al., “Data from lidc-idri. the cancer imaging archive,” *DOI [http://doi.org/10.7937/K9\(7\)2015](http://doi.org/10.7937/K9(7)2015)* (2015).
- [8] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P.-Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Vande Casteele, A., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., and Clarke, L. P., “The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans,” *Medical Physics* **38**(2), 915–931 (2011).
- [9] Jaccard, P., “The distribution of the flora in the alpine zone.1,” *New Phytologist* **11**(2), 37–50 (1912).