BOOSTING SEGMENTATION ACCURACY OF THE DEEP LEARNING MODELS BASED ON THE SYNTHETIC DATA GENERATION

V. V. Danilov[1,5,*], O. M. Gerget[1], D. Yu. Kolpashchikov[1], N. V. Laptev[1], R. A. Manakov[2],
L. A. Hérnandez-Gómez[3], F. Alvarez[4], M. J. Ledesma-Carbayo[5]

[1] Research Laboratory for Processing and Analysis of Big Data, Tomsk Polytechnic University,
Tomsk, Russia - viacheslav.v.danilov@gmail.com
[2] Department of Software Engineering, Research Laboratory "Gamma Technologies",
Almaty, Kazakhstan - ram290495@gmail.com
[3] Department of Signals, Systems and Radiocommunications, Technical University of Madrid,
Madrid, Spain - luisalfonso.hernandez@upm.es
[4] Group of Application of Visual Telecommunications, Technical University of Madrid,
Madrid, Spain - federico.alvarez@upm.es
[5] Biomedical Image Technologies Group, Technical University of Madrid,
Madrid, Spain - mariajesus.ledesma@upm.es

Commission II, WG II/5

KEY WORDS: Data Synthesis, Echocardiography, Catheter Segmentation, Forward Kinematics, Spline Coordinate System

ABSTRACT:

In the era of data-driven machine learning algorithms, data represents a new oil. The application of machine learning algorithms shows they need large heterogeneous datasets that crucially are correctly labeled. However, data collection and its labeling are time-consuming and labor-intensive processes. A particular task we solve using machine learning is related to the segmentation of medical devices in echocardiographic images during minimally invasive surgery. However, the lack of data motivated us to develop an algorithm generating synthetic samples based on real datasets. The concept of this algorithm is to place a medical device (catheter) in an empty cavity of an anatomical structure, for example, in a heart chamber, and then transform it. To create random transformations of the catheter, the algorithm uses a coordinate system that uniquely identifies each point regardless of the bend and the shape of the object. It is proposed to take a cylindrical coordinate system as a basis, modifying it by replacing the Z-axis with a spline along which the h-coordinate is measured. Having used the proposed algorithm, we generated new images with the catheter inserted into different heart cavities while varying its location and shape. Afterward, we compared the results of deep neural networks trained on the datasets comprised of real and synthetic data. The network trained on both real and synthetic datasets performed more accurate segmentation than the model trained only on real data. For instance, modified U-net trained on combined datasets performed segmentation with the Dice similarity coefficient of $92.6\pm2.2\%$, while the same model trained only on real samples achieved the level of $86.5\pm3.6\%$. Using a synthetic dataset allowed decreasing the accuracy spread and improving the generalization of the model. It is worth noting that the proposed algorithm allows reducing subjectivity, minimizing the labeling routine, increasing the number of samples, and improving the heterogeneity.

1. INTRODUCTION

Many machine learning algorithms are fairly sensitive to the datasets used for training. Therefore, it is critical to have access to high-quality datasets. Typically, training and test samples come from the same statistical distribution. Whilst the paucity of flexible and rich enough datasets limits the ability of machine learning or statistical modeling techniques and leaves the algorithm generalization capability superficial. However, if it is possible to generate labeled samples with a distribution close enough to the studied one, these samples can be used to test solution performance and reliability. Synthetic datasets that are generated programmatically can help immensely in the field of machine learning. These datasets are not collected by any real-life survey or experiment. Their main purpose, therefore, is to be flexible and rich enough to help in conducting experiments with various classification, segmentation, and object detection algorithms.

Nowadays data synthesis algorithms are quite popular in the

---

* Corresponding author

healthcare industry. For instance, K. Antczak and Ł. Liberadzki try to solve a stenosis detection problem based on convolution neural networks (Antczak, Liberadzki, 2018). To increase the training dataset, the authors use a relatively straightforward algorithm generating artificial patches. To draw veins with stenosis and atherosclerotic plaques, the algorithm uses the Bézier curves. The classifier used in this study is trained in two stages. In the first stage, it uses artificial patches for training and then tuned-up on real samples. Such an approach allowed the authors to reach a classification accuracy equal to 0.90.

Another approach connected with the Generative Adversarial Networks (GANs) is widely used for medical image synthesis (Yi et al., 2018). One of the studies, performed by J.T. Guibas et al., describes an implementation of dual GANs for medical image synthesis (Guibas et al., 2017). The first GAN is used for the generation of a segmentation mask. The second GAN translates the masks produced by the first GAN to photorealistic images. To tackle the problem of image synthesis, P. Costa et al. developed a method that learns to synthesize eye fundus images directly from data (Costa et al., 2017). In this method,

adversarial networks and adversarial autoencoders are used to synthesize retinal images. The authors pair real retinal images with their respective vessel trees by means of a vessel segmentation technique. Then these pairs are used to learn a mapping from a binary vessel tree to a new retinal image. The produced data can help generate labeled data for training and testing the models dedicated to retinal image analysis. It should be also noted that GANs are also used for cross-modality synthesis. The latter allows generating new training samples with the appearance constrained by the anatomical structures delineated for the available modality.

Deep learning methods require extensive and representative samples of data to enable high-quality training of neural networks. However, acquiring such data is sometimes very difficult or even impossible especially when experimenting and labeling are expensive. When solving the problem of localization and segmentation of the distal end of the catheter inside the heart, we encountered the problem of insufficient data and weak representativeness. To solve this problem we propose a new algorithm for synthesizing echocardiography data with inserted medical devices. The key idea behind the proposed algorithm is to insert a catheter from one three-dimensional image to another with the ability to control its position and shape. To maintain an accurate configuration of the inserted catheter, the algorithm uses the kinematics of continuum robots. Despite there are such strategies dealing with homogeneous and small datasets as augmentation and GANs, they have a common weakness, relying on an initial dataset. As a result, all generated elements are related to real ones, which can be a significant obstacle if the dataset is small enough. Having used the proposed algorithm, we overcame this problem since the algorithm allowed generating as many new configurations and positions of the catheter as possible. It is also important to note that existing methods of data augmentation do not consider the correlation of neighboring slices/images and randomly apply a particular transformation. In turn, the proposed algorithm is designed to transform the object of study, taking into account the relationship of data.

2. SOURCE DATA

The initial data were obtained by means of epicardial three-dimensional echocardiography when performing cardiac surgery on three Yorkshire porcine hearts. This dataset was collected at Boston Children's Hospital (Boston, USA). During each surgery, a medical instrument (catheter) was inserted into the cavity of the left ventricle. The transthoracic X7-2t sensor was placed on the epicardium of the left ventricle apex. In addition to the transthoracic sensor, the Philips iE33 ultrasound machine and PMS5.1 ultrasound software were used to acquire the data. In the process of data collection, we acquired 75 three-dimensional ultrasound grayscale samples of 176x176x208 voxels each. Some of these samples are reflected in Fig. 1. It is worth noting that the catheter is poorly visible to the human eye on the data of echocardiography. In this regard, we highlighted the catheter in green circles and ellipses.

Additionally, we obtained data with empty cavities of human hearts where no medical instrument was used. This dataset was collected at the Cardiology Research Institute (Tomsk, Russia). The total amount of data with empty heart cavities made up 600 3D images. Examples of such images are shown in Fig. 2. Further in our study, these images are used for placing a catheter into their empty cavities.
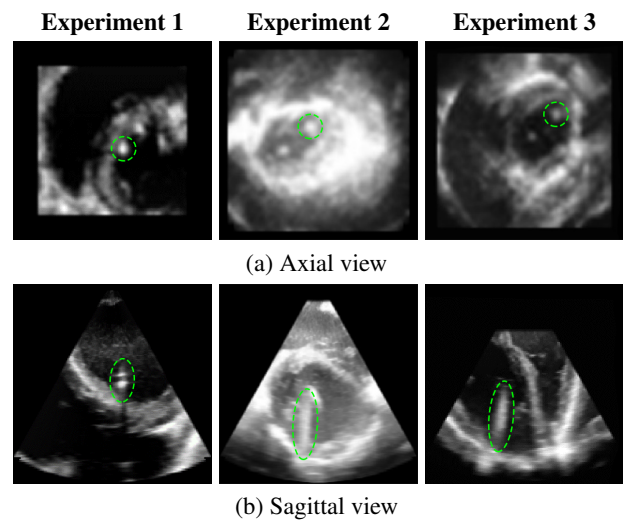
Experiment 1 **Experiment 2** **Experiment 3**

(a) Axial view

(b) Sagittal view

Figure 1. Examples of the source three-dimensional dataset with a catheter inserted into the left ventricle cavity

Experiment 1 **Experiment 2** **Experiment 3**
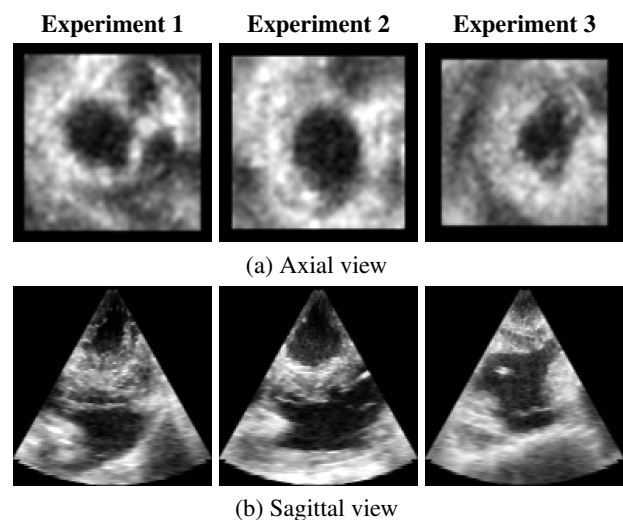

(a) Axial view

(b) Sagittal view

Figure 2. Examples of the source three-dimensional dataset with empty heart cavities

3. METHODS

3.1 Data synthesis

The application of neural networks for the localization and segmentation of a medical instrument in three-dimensional echocardiography requires a relatively large training dataset. The lack of such images in sufficient quantities is one of the key problems of the deep learning approach. One of the solutions to this problem is the generation of new artificial images based on existing ones. The concept of this generation is the distortion of the image with a medical instrument and its transfer to the real image of the heart with empty cavities. The proposed data synthesis algorithm uses different transformations such as bending, twisting, scaling and displacement. Thus, to generate artificial three-dimensional echocardiography of the heart containing the distal end of the catheter, the following inputs are needed: a three-dimensional image of the catheter, a three-dimensional image of the heart with empty cavities, the starting point and orientation of the catheter. The result of the generator is three-dimensional images containing a catheter inside the anatomical structures of the heart. The generator model has
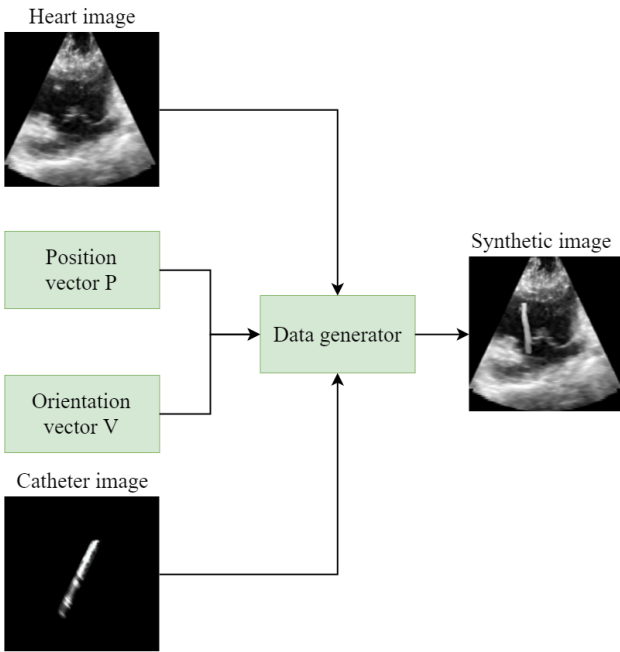
Figure 3. Scheme of the synthetic data generator

4 inputs and is shown in Fig. 3. It is worth adding that a pair of vectors $(\overline{P}, \overline{V})$ sets the position and orientation of the catheter in a new blank image.

To implement the transfer and transformation of the catheter, we developed our own coordinate system based on a cubic spline. This system allows working flexibly with the points of the catheter and carrying out all the necessary transformations. The spline passes through the longitudinal axis of the catheter and sets its configuration. In turn, all points of the catheter are calculated relative to the spline using (ρ, φ, z) coordinates according to the principle of a cylindrical coordinate system. The spline plays the role of the Z-axis in the proposed system. While the Z-axis is represented by a straight line in the classical cylindrical coordinate system. Thus, the z coordinate is the length of the spline from its beginning to the point O on the spline (see Fig. 4). At this point, the X-axis is plotted; the angle and distance to the point M belonging to the catheter are determined. The point O is the closest point on the spline relative to the point M, and the segment OM is perpendicular to the tangent line of the spline at the point O.

This approach allows us to position the points of the catheter relative to its axis, regardless of the shape of the spline. The transfer of a point from the spline occurs according to the following formula:
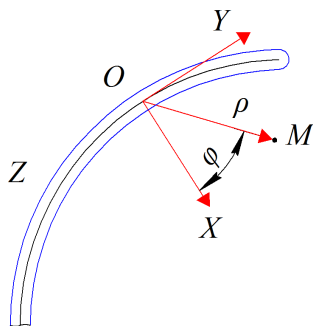
$$M^{'} = F^{-1}\left(F\left(M, \ S_1\right), \ S_2\right) \qquad (1)$$

where F is the transformation described above, F^{-1} is the inverse transformation. This expression allows mapping points from spline S_1 to spline S_2, which solves the problem of transforming the overall shape of the catheter. Additionally, the catheter can be stretched by normalizing the z coordinates to the length of the spline using the following transformation:

$$z^{'} = \frac{z}{l_1} l_2 \qquad (2)$$

where l_1 and l_2 are the lengths of the corresponding splines S_1 and S_2, respectively.

It is impossible to unequivocally plot the X-axis at the point O without additional information about its direction. This is because the condition of perpendicularity to the tangent at a given point is satisfied by an infinite set of vectors lying in the plane perpendicular to the tangent. For the exact construction of this axis, it is proposed to use the function $D(z)$, which determines the direction of the axis at the nodes of the main spline, and is interpolated between the nodes. However, the interpolation of this function does not guarantee the perpendicularity of the tangent vector at points other than nodal. This vector is corrected so that the condition of perpendicularity to the tangent is met with minimal deviations. Thus, the X-axis can be found using the following formulas:

$$\alpha = \frac{-(\bar{K}(z), \bar{D}(z))}{(\bar{K}(z), \bar{K}(z))} \qquad (3)$$

$$\bar{X} = \bar{D}(z) + a\bar{K}(z) \qquad (4)$$

where $\bar{K}(z)$ is the tangent vector, $\bar{D}(z)$ is the function that sets the twisting of the catheter around its axis, which is one of the ways to transform the data.

The catheter configuration is generated using forward kinematics algorithms, which build an axial spline, as well as the vector function $\bar{D}(z)$ for the given bending angles and the lengths of the corresponding catheter joints. The kinematics algorithm of the catheter is described in more detail in (Kolpashchikov et al., 2018).

The mapping process occurs for each voxel separately. It is worth noting that in the general case the result of the mapping is the real Cartesian coordinates, while the values of the three-dimensional voxels are determined only for integer coordinate values. In this regard, instead of a rounding method, we propose to use a trilinear interpolation on a three-dimensional regular grid, which improves image quality.

The generation of a new configuration is randomly carried out until an obtained configuration completely fits into the required anatomical structure of the heart. The control is performed by checking the catheter point cloud and the mask of a blank three-dimensional image, where the catheter is placed. It should be also noted that the position and orientation changes of the catheter, placed inside the heart chambers, are randomly generated. The results of artificial data generation are reflected below in Section 4.1.

Figure 4. Definition of the coordinates of the point M

3.2 Deep learning

To estimate how synthetic data affect the model performing segmentation, we used an encoder-decoder U-net architecture (Ronneberger et al., 2015). However, when implementing the original architecture without normalization, the gradient descent algorithm did not converge well, so that the segmentation accuracy did not exceed the level of 5%. Therefore, we decided to make several modifications in the architecture, where the encoder and decoder blocks were significantly reworked. Most layers were replaced and the proposed modified U-net architecture contained dilated convolution layers, instance normalization, ELU activation layers, max-pooling layers, and transposed convolution layers. The proposed modification of the U-net architecture is shown in Fig. 5.

The introduction of the dilated convolution (Yu, Koltun, 2015) into this study is connected with minimizing the number of trained weights. It is easy to note that when using the dilation rate $l > 1$, the number of trainable weights in comparison with the regular convolution is significantly reduced while the size of the receptive window remains unchanged. For example, a standard 5×5 convolution filter has 25 trainable weights. In turn, a dilated convolution with the same filter size has 9 non-zero weights. This modification allows using filters of a larger size, which, in turn, increases the field of view of the convolution kernel.

The use of ELU activation layers accelerates the learning process, partially eliminates the problem of vanishing gradients, and also increases the classification accuracy of neural networks (Clevert et al., 2016). In contrast to the ReLU activation function, the ELU activation function has a non-zero negative component. The use of negative gradients makes it possible to shift the mean activation value to zero, which, in turn, helps to minimize unnecessary shifts and offsets. A similar procedure is performed by batch normalization. However, the ELU activation layer performs this task with less computational complexity.

3.3 Experiments description

To evaluate the influence of synthetic data on the accuracy and generalization ability of neural networks, 4 models were trained based on the proposed modified U-net architecture, described in Section 3.2. Initially, we trained one model using the original non-synthetic dataset. As indicated in Section 2, this dataset was obtained from three porcine hearts with catheters inserted into the left ventricles for surgical purposes. Having applied the proposed algorithm, we generated synthetic samples with the catheters inserted into the echocardiographic images of empty human hearts. Once the synthetic dataset was generated, we gradually added synthetic samples to the training dataset. By performing these experiments, we checked how synthetic data influence the accuracy of neural networks, and whether it brings positive or negative dynamics. In total, we performed four different experiments varying Real Data Ratio (*RDR*), which is calculated as follows:

$$RDR = \frac{Real\ samples}{Real\ samples + Synthetic\ samples} \quad (5)$$

A short description of the data used in the experiments is presented below in Table 1.

To estimate the segmentation accuracy with different *RDR* values, the Dice similarity coefficient (*DSC*) was used as the main

Table 1. Data distribution within 4 different experiments

Experiment	Real samples	Synthetic samples	Samples in total	RDR
1	75	0	75	1.00
2	75	75	150	0.50
3	75	150	225	0.33
4	75	225	300	0.25

segmentation metric. This metric is computed as follows:

$$DSC = 2 \cdot \frac{|A \cap B|}{|A| + |B|} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (6)$$

where $|A|$ and $|B|$ are the cardinalities of set A and B, TP is the number of true positives, FP is the number of false positives, FN is the number of false negatives.

4. RESULTS

This section presents the results obtained by the proposed algorithm for data synthesis. In addition to visualizing synthesis and segmentation results, we reflect accuracy assessment. In Section 4.2, we demonstrate *DSC* distributions varying the different *RDR* values.

4.1 Synthesis and segmentation

Having performed the proposed algorithm on the source data with empty cavities, we generated 225 three-dimensional samples of echocardiography with the catheter inserted into them. One of these samples is reflected in Fig. 6, where the catheter is shown in green circles and ellipses. As shown, the catheter was placed in the left ventricle according to the constraints of this cavity.

Once the real and synthetic datasets were obtained, the modified U-net was trained with different values of *RDR*. In total, 4 models were trained. An example of segmentation of a 3D image by the modified U-net is shown in Fig. 7. As seen, the proposed modification of U-net segment the catheter accurately.

4.2 Accuracy assessment

Having the ground truth of the data, we performed an analysis of segmentation accuracy. According to the obtained results, we observed an inverse proportionality between *RDR* and *DSC* i.e. the less the *RDR* is, the more the *DSC* is. The results of the calculated *DSC* are shown in Table 2. Additionally, we compared the *DSC* distributions obtained with different *RDR* values. This comparison is reflected in Appendix A. The *DSC* distribution at *RDR* = 1.00 was considered as a baseline distribution which means that the network was trained and tested only on real data. The remaining distributions obtained at *RDR* = 0.50, *RDR* = 0.33 and *RDR* = 0.25 were compared to the baseline distribution.

According to the *DSC* distirbution comparsion, the average segmentation accuracy of the modified U-net increases with decreasing *RDR* metric (see Appendix B). Nevertheless, the average segmentation accuracy of the network makes up approximately 90%. It is also worth noting that there is no clear asymptotic saturation. Therefore, mixing data, for example, at a ratio of 1:4, can presumably lead to either an increase in the *DSC* or to the achievement of its asymptote.
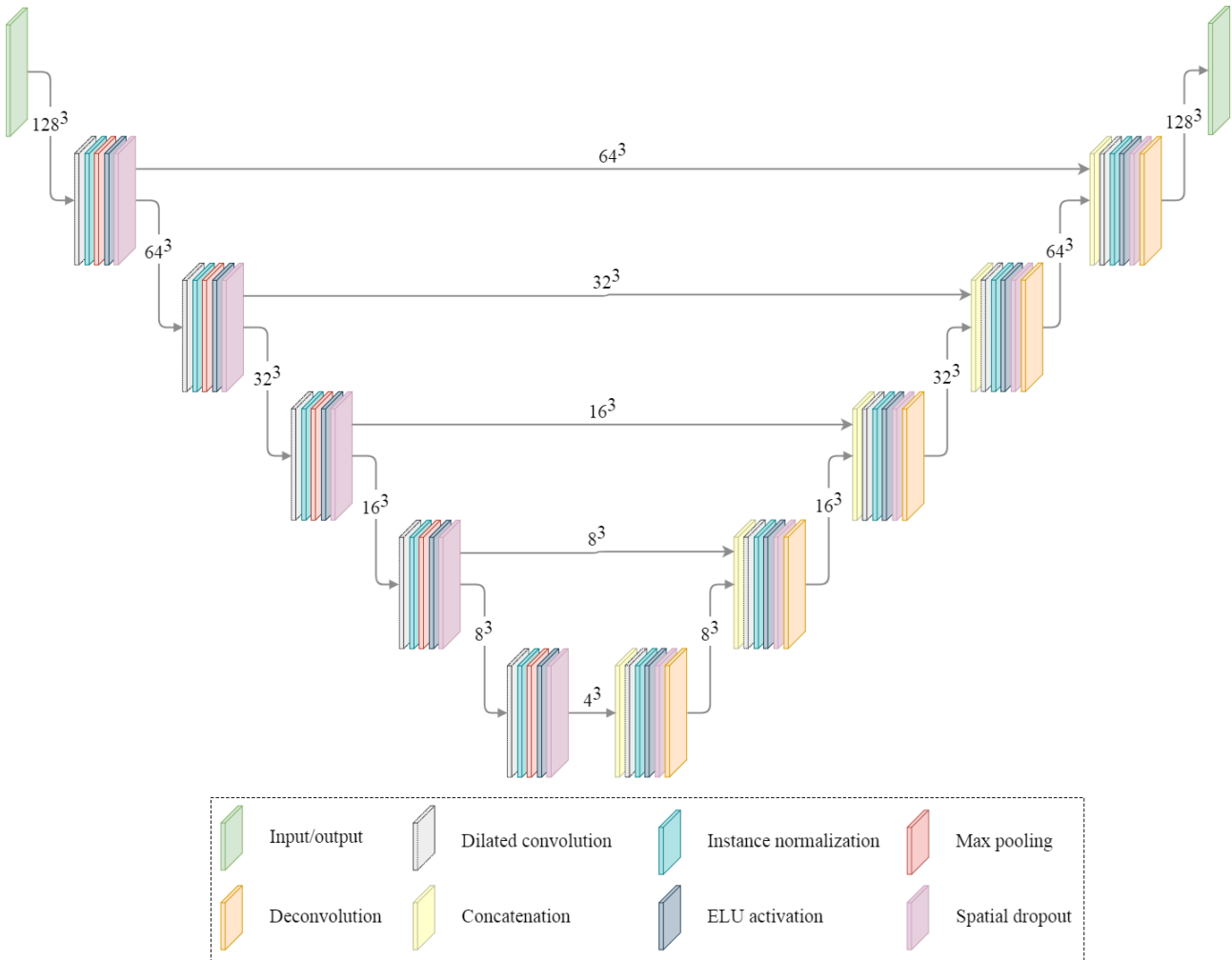
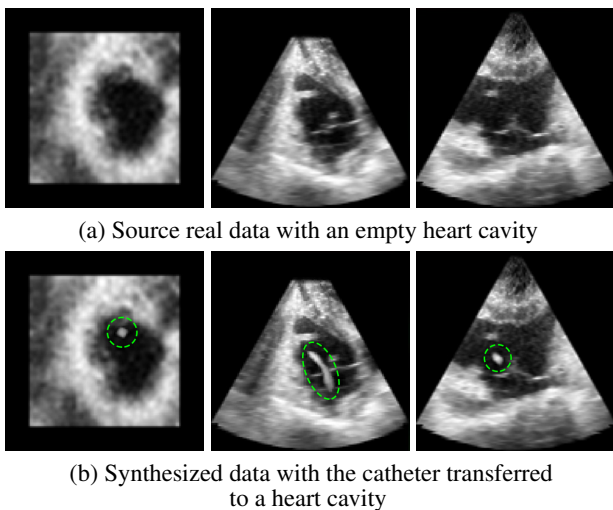Figure 5. Modified U-net architecture used for concept proof.

(a) Source real data with an empty heart cavity

(b) Synthesized data with the catheter transferred to a heart cavity

Figure 6. An example of an artificial sample synthesized by the the proposed algorithm.

(a) Source data with the ground truth mask (red)

(b) Segmentation performed by the modified U-net (white) and the ground truth (red)
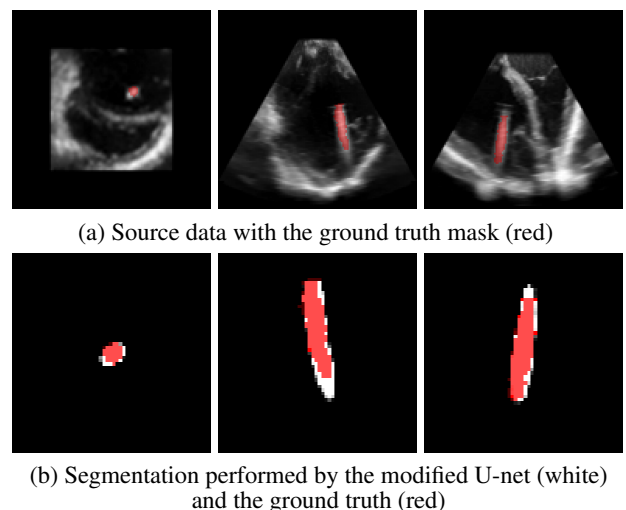
Figure 7. Segmentation of the catheter based on the dataset with $RDR = 0.25$.

Table 2. *DSC* of the modified U-net with different *RDR* values

	RDR			
	1.00	**0.50**	**0.33**	**0.25**
Train	0.85±0.05	0.87±0.03	0.88±0.05	0.93±0.01
Val	0.88±0.03	0.88±0.03	0.89±0.04	0.93±0.02
Test	0.86±0.04	0.88±0.03	0.90±0.03	0.93±0.02

5. DISCUSSION

Despite the fact that the proposed algorithm helped us to successfully solve the problem of catheter segmentation in three-dimensional echocardiographic images, it has several limitations. The first limitation is related to the object shape. The current version of the algorithm is only applicable cylinder-shaped objects. The second limitation of the algorithm does not accurately take into account ultrasound effects i.e. noise, structure, texture, etc. This drawback is partially solved by the trilinear interpolation used in the algorithm. In order to completely solve this issue, some image processing or deep learning techniques can be applied. It should be also noted that synthetic data generation is a relatively time-consuming process. On average 54 seconds are needed to generate one three-dimensional synthetic image of $128\times128\times128$ voxels. This is due to the relatively lengthy procedure of integrating the spline function along the catheter length, which is performed for each voxel. However, data generation does not have to be run in real-time. It is worth noticing that the usage of real input data allows obtaining images with a high degree of plausibility. In turn, forward kinematics allows simulating the shape of the real catheter.

One of the requirements applied to data synthesis is the ability of the synthesizer to generate data easier than it can be acquired in real life. It should be noted that creating a data synthesis algorithm may be very time-consuming. However, if the cost of creating this algorithm is lower than the cost of collecting a training set of real data, it is better to lean towards data synthesis. An important restriction of the data synthesizer is its ability to generate a distribution that is close to a set of real distribution. Another restriction is related to the randomness of data synthesizing. It means that the underlying random process should be precisely controlled and tuned. As an additional feature, data synthesis algorithms should apply random noise to an image in a controllable manner.

6. CONCLUSION

When using machine learning to solve a segmentation or localization task, a dataset should be large and representative. If the latter fails, the network may have a weak generalizing ability. To solve the problem of data unrepresentativeness, we proposed an algorithm inserting and transforming a cylindrical object into a constrained area. The proposed algorithm was used for the generation of synthetic data of a catheter located inside the cavities of the heart. In order to control the correct shape of the catheter, we applied forward kinematics of the real catheter. As for the catheter insertion area and its constraints, the image where the catheter is inserted should have a labeled mask. The latter is used to control the placement of the catheter inside the anatomical structure of the heart. Having generated the data, we checked how the proposed modification of U-net performed

segmentation. According to the obtained results, we observed a positive dynamics for the models used both real and synthetic data. For instance, modified U-net performed segmentation of the catheter with a *DSC* of 92.6±2.2% for *RDR* = 0.25 and 86.5±3.6% for *RDR* = 1.0.

REFERENCES

Antczak, K., Liberadzki, Ł., 2018. Stenosis Detection with Deep Convolutional Neural Networks. *MATEC Web of Conferences*, 210, 04001. https://www.matec-conferences.org/10.1051/matecconf/201821004001.

Clevert, D.-A., Unterthiner, T., Hochreiter, S., 2016. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *Conference proceedings at ICLR 2016*, 1–14.

Costa, P., Galdran, A., Meyer, M. I., Mendonça, A. M., Campilho, A., 2017. Adversarial synthesis of retinal images from vessel trees. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10317 LNCS, 516–523.

Guibas, J. T., Virdi, T. S., Li, P. S., 2017. Synthetic Medical Images from Dual Generative Adversarial Networks. http://arxiv.org/abs/1709.01872.

Kolpashchikov, D., Laptev, N., Danilov, V., Skirnevskiy, I., Manakov, R., Gerget, O., 2018. FABRIK-Based Inverse Kinematics For Multi-Section Continuum Robots. *Proceedings of the 2018 18th International Conference on Mechatronics*, IEEE, Brno, Czech Republic, 1–8.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. 234–241. http://link.springer.com/10.1007/978-3-319-24574-4_28 http://arxiv.org/abs/1505.04597.

Yi, X., Walia, E., Babyn, P., 2018. Generative Adversarial Network in Medical Imaging: A Review. *Medical Image Analysis*, 58, 101552. https://linkinghub.elsevier.com/retrieve/pii/S1361841518308430 http://arxiv.org/abs/1809.07294 http://dx.doi.org/10.1016/j.media.2019.101552.

Yu, F., Koltun, V., 2015. Multi-Scale Context Aggregation by Dilated Convolutions. *Proceedings at ICLR 2016*.
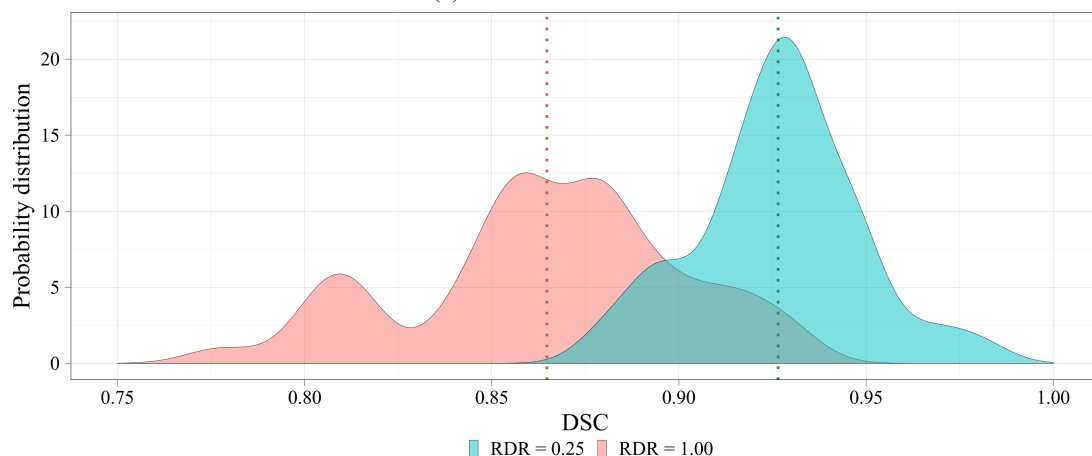
APPENDIX A

Comparison of the *DSC* distributions for different *RDR* values
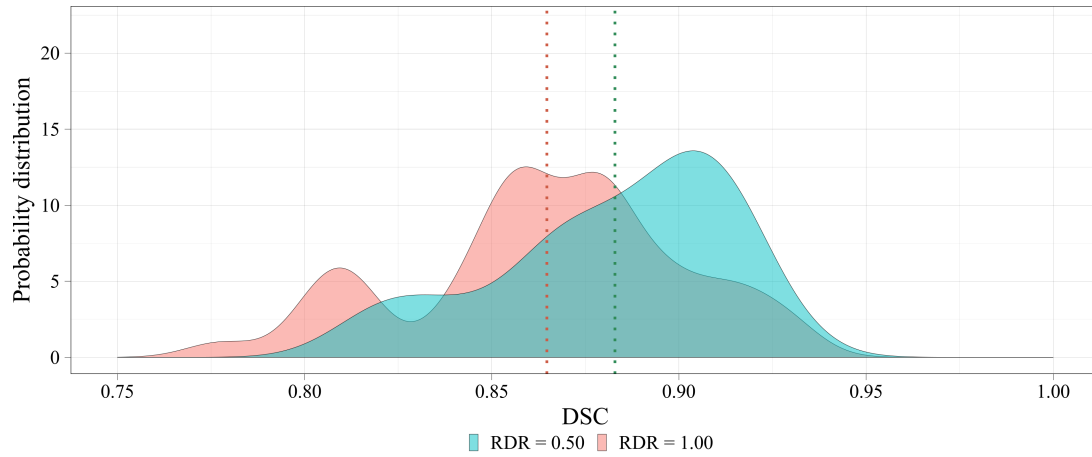
(a) *RDR* = 0.50 vs *RDR* = 1.00
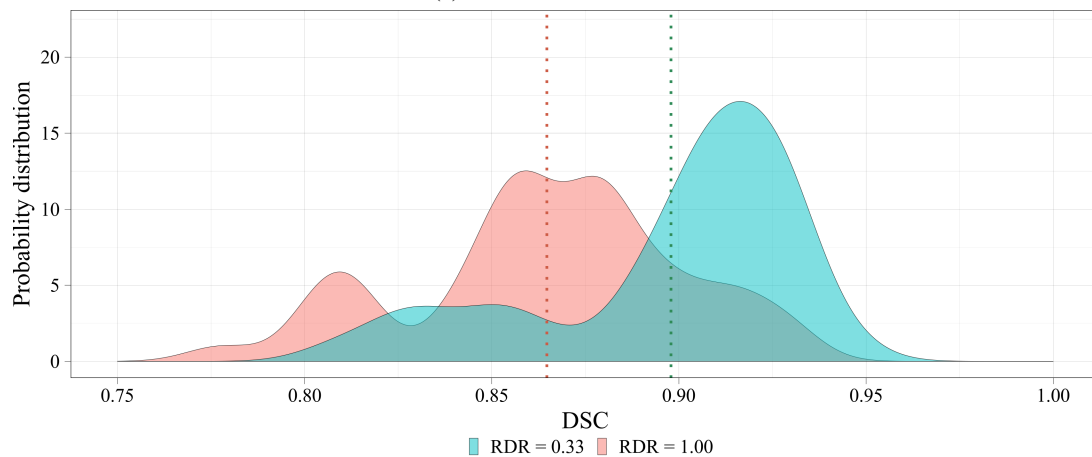
(b) *RDR* = 0.33 vs *RDR* = 1.00

(c) *RDR* = 0.25 vs *RDR* = 1.00

APPENDIX B

Comparison of *DSC* for different *RDR* values over training, validation, and testing subsets

(a) Training (b) Validation (c) Testing