OPINION

Deep Learning in Radiology: Now the Real Work Begins

Carolina Lugo-Fagundo, BS, Bert Vogelstein, MD, Alan Yuille, PhD, Elliot K. Fishman, MD

INTRODUCTION

Machine learning systems build statistically based mathematical models that program computers to optimize a performance criterion from sample data [1]. Deep learning is a form of machine learning that attempts to mimic the interactivity of layers of neurons in the neocortex, the part of the cerebral cortex where most higher thinking occurs [2]. Deep learning methods can now be used to recognize patterns in digital representations of sound and images as well as in other data types.

Recognizing its power, medical researchers and scientists have been exploring the role of deep learning methods to create tools to alert clinicians to focus on outliers identified in data sets and as aids for earlier and more accurate diagnosis. Recent studies have highlighted the potential power of deep learning in a variety of medical applications such as imaging-guided survival time prediction of patients with brain tumors [3], bladder cancer treatment response assessment [4], lung nodule [5] and mammographic classifications [6], diagnosis of breast lesion pathologies [7], and body part recognition [8]. However, many studies fail to emphasize the importance of the data collection required to build the algorithms.

In general, the scientific community agrees that large volumes of data are required for building robust deep learning algorithms [9,10]. Most recently, Esteva et al developed an algorithm to separate gross images of skin lesions into nonproliferative benign or malignant lesions [9]. The investigators utilized a set of 129,450 images of skin lesions for computer training; 2,032 of those images were of skin diseases diagnosed using noninvasive visual analysis or biopsy testing by dermatologists [9]. Esteva et al specified that the machine's performance can be enhanced if it is trained with a larger data set. Similarly, for deep learning to be successful in radiology, it will be necessary to collect and annotate potentially thousands of cases representing a range of pathologies.

These detailed data are used to create deep networks that are trained in a supervised way, with a method that uses more knowledge and information. We have previously proven that weak supervision, which specifies a bounding box around an object instead of per pixel labeling, does not work for deep network building of complex data [11]. This method might function in simpler tasks like recognizing a dog from a cat, but not for compound data like pancreatic anatomy and pathology.

THE FELIX PROJECT

The Felix Project, funded by the Lunstgarten Foundation, is a multidisciplinary research study designed to determine if deep learning algorithms can be developed to aid in the interpretation of CT and MR images of the pancreas. The ultimate goal is to develop a tool to reduce radiologist errors and to diagnose changes in the pancreas that can signal early stage, curable pancreatic cancers. The Felix team includes experts in medical imaging, pathology, oncology, and computer science, including experts with experience in artificial intelligence and deep learning.

The first step of the project is to carefully annotate a large set of normal images for machine learning training. Once the computer is trained on the normal CT data sets, the computer should be able to identify and recognize a normal pancreas. The next step will be to train the computer to identify an abnormal pancreas by providing it

1546-1440/17/$36.00 ■ http://dx.doi.org/10.1016/j.jacr.2017.08.007

with images of patients with a wide range of pancreatic pathologies, including adenocarcinomas, neuroendocrine tumors, and pancreatic cysts of various types.

DEEP LEARNING, DEEP WORK

If we have learned anything from the initial phase of our project, it is this: deep learning requires deep work. The algorithms developed using deep learning are only as good as the training sets used to train the system. In the case of CTs, this requires the careful annotation of hundreds or even thousands of images in both the arterial and venous phases. Each CT image we provide shows 21 distinct (Fig. 1) segmentations, including segmentation of the pancreas as well as other organs and vessels adjacent to the pancreas. The creation of well-annotated training sets for this study is extremely labor intensive. Our initial goal is to segment at least 1,000 venous and arterial CT images of normal pancreas, and to date we have annotated over 800 cases and are using them as the ground truth to create our complex algorithms. The veracity of the ground truth is critical because the deep learning algorithms can only be as truthful as this reference [9].

Segmentation of the CT images requires accuracy and precision. Using state-of-the-art segmentation software Velocity (Varian Medical System, Atlanta, Georgia, USA) (Fig. 2) and a four-member segmentation team, each data set takes about 3 hours to segment. Each organ is manually contoured using 0.75-mm CT images, making sure that the margins are clearly defined and differentiated from the other structures. Even though the segmentation software offers edge-detection tools and a range of texture-sensitive tools, contouring still requires tedious manual input.

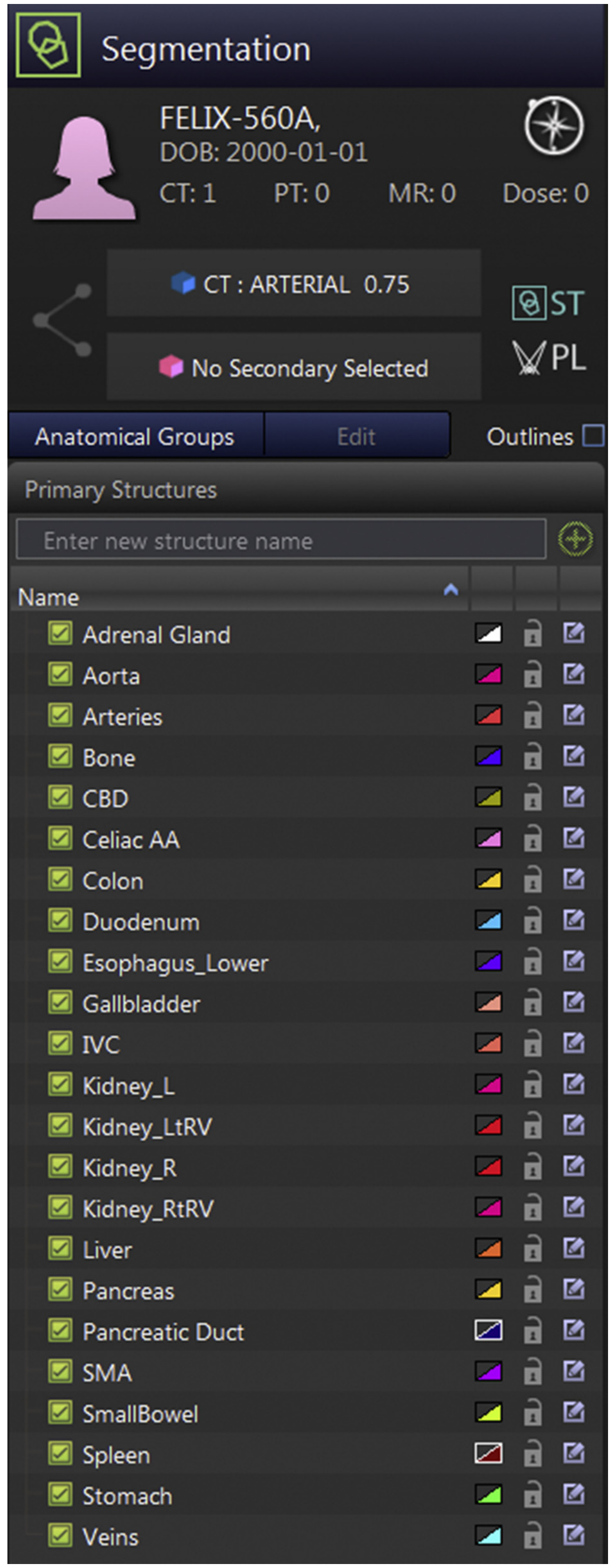

Fig 1. List of the 21 separate organs and vessels that are contoured.

To streamline the process and minimize variations in contouring, we developed best practice

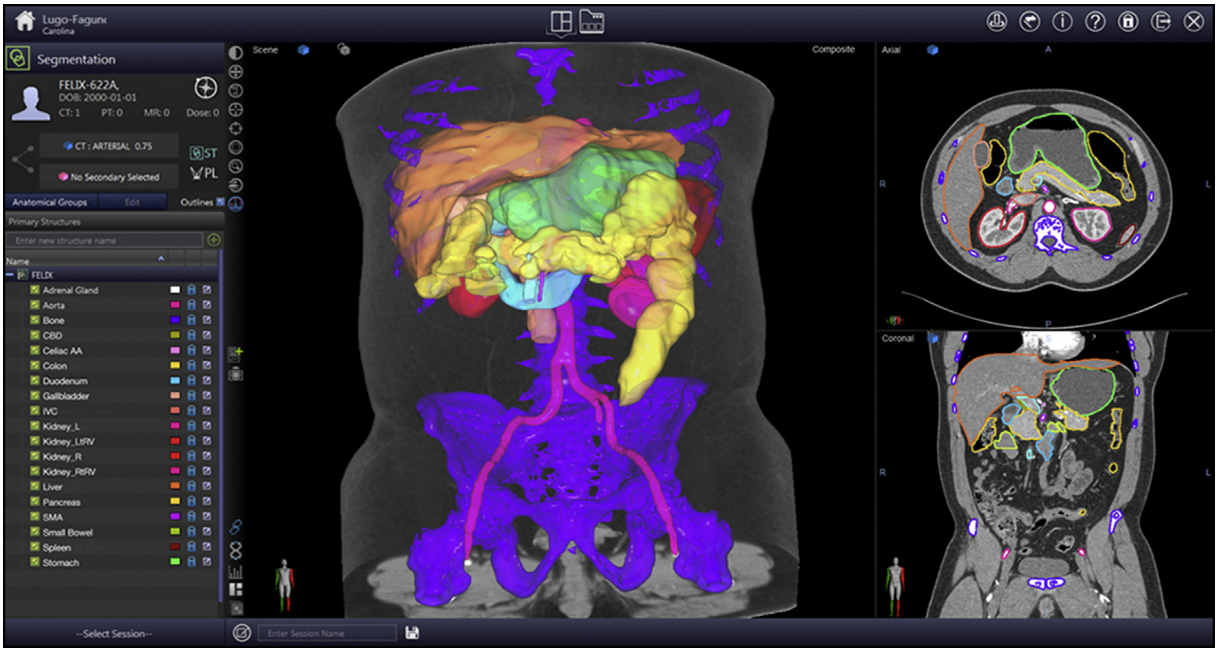

Fig 2. Schematic representation of the segmentation work screen. It contains 3-D axial and coronal views of an abdominal 0.75-mm CT in the arterial phase.

guidelines and optimization techniques that can be updated if needed. This segmentation process is also expensive, considering the cost of salaries, segmentation software, and training. Thus a considerable amount of time, resources, expertise, and labor are necessary for a project like this, even before the creation of the deep learning computer algorithms.

CONCLUSION

The quantity and quality of the training set are critically important in the development of state-of-the-art deep learning in radiology. One way to expedite the process would be to have a number of institutions develop standard processes and best practice guidelines for data segmentations and share these data sets. These collaborations would not limit algorithm development, which is the ultimate goal, but speed it along by increasing the number of normal and abnormal data sets available. Finally, it is likely that these algorithms will undergo continuous learning and optimization as they are applied in clinical practice in much the same way voice recognition algorithms improve from their real-world deployment. With the extensive amount of work that lies ahead, Winston Churchill's statement seems appropriate: This "is not even the beginning of the end. But it is, perhaps, the end of the beginning" [12].

ACKNOWLEDGMENTS

The authors acknowledge the following people who contributed to this article: Ralph Hruban, MD, Karen Horton, MD, Linda Chu, MD, and Kenneth Kinzler, MD.

REFERENCES

1. Alpaydın E. Introduction to machine learning. Cambridge, MA: Massachusetts Institute of Technology Press; 2014.
2. Hof RD. Deep learning: with massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. MIT Technology Review 2013;116:32.
3. Nie D, Zhang H, Adeli E, Liu L, Shen D. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. Medical image computing and computer-assisted intervention. Medical Image Computing and Computer-Assisted Intervention 2016 Lecture Notes in Computer Science, Athens, Greece 2016:212-20.
4. Cheng JZ, Ni D, Chou YH, et al. Bladder cancer segmentation in CT for treatment response assessment: application of deep-learning convolution neural network—a pilot study. Tomography 2016;2:421-9.
5. Wang C, Elazab A, Wu J, Hu Q. Lung nodule classification using deep feature fusion in chest radiography. Comput Med Imaging Graph 2016;57:10-8.
6. Kooi T, Litjens G, Ginneken BV, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017;35:303-12.
7. Cheng J-Z, Ni D, Chou Y-H, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and

pulmonary nodules in CT scans. Sci Rep 2016;6:24454.

8. Yan Z, Zhan Y, Peng Z, et al. Bodypart recognition using multi-stage deep learning. Lecture notes in computer science information processing in medical imaging. Inf Process Med Imaging, Springer, New York. 2015:449-61.

9. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-8.

10. Jha S, Topol EJ. Adapting to artificial intelligence. JAMA 2016;316:2353.

11. Papandreou G, Chen LC, Murphy K, Yuille AL. Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation. The IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, Washington DC 2015:1742-50.

12. Knowles, E. (2014). Oxford dictionary of quotations (5th ed.). Oxford: Oxford University Press.

Elliot K. Fishman, MD and Carolina Lugo-Fagundo, BS, are from The Russell H. Morgan Department of Radiology and Radiologic Science, Johns Hopkins School of Medicine, Baltimore, Maryland. Bert Vogelstein, MD, is from the Department of Molecular Biology and Genetics. Alan Yuille, PhD is from the Department of Cognitive Science, Johns Hopkins School of Medicine, Baltimore, Maryland.

The Lustgarten Foundation, Myriad Genetics, miDiagnostics, PapGene, Personal Genome Diagnostics, Morphotek, Syxmex Inostics, Exelixis GP. The Felix project is funded in part by a grant from The Lustgarten Foundation. The authors have no conflicts of interest related to the material discussed in this article.

Elliot K. Fishman, MD: The Russell H. Morgan Department of Radiology and Radiologic Science, Johns Hopkins School of Medicine, 601 N Caroline Street, Room 3254, Baltimore, MD 21287; e-mail: efishman@jhmi.edu.