Promises and limitations of deep learning for medical image segmentation

Christian S. Perone, Julien Cohen-Adad

NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada

Correspondence to: Christian S. Perone. NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC H3T 1J4, Canada. Email: christian.perone@gmail.com.

Comment on: Minnema J, van Eijnatten M, Kouw W, *et al.* CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. Comput Biol Med 2018;103:130-9.

Received: 26 December 2018; Accepted: 08 January 2019; Published: 10 January 2019.

doi: 10.21037/jmai.2019.01.01

View this article at: http://dx.doi.org/10.21037/jmai.2019.01.01

It is not a secret that recent advances in deep learning (1) methods have achieved a scientific and engineering milestone in many different fields such as natural language processing, computer vision, speech recognition, object detection, and segmentation, to name a few. Different applications of deep learning to medical imaging started to appear first in workshops, conferences and then in journals. According to a recent survey (2), the number of papers grew rapidly in 2015 and 2016. Nowadays, deep learning methods are pervasive throughout the entire medical imaging community, with Convolutional Neural Networks (CNNs) being the most used model for tasks such as dense prediction (or segmentation), detection and classification. In the same survey, which analyzed more than 300 contributions in the field, the authors found that computed tomography (CT) was the third most used imaging modality.

Deep learning methods for tasks that require a dense output such as voxel-wise classification (which is common in image segmentation), have improved by many orders of magnitude in the past years (3). They are slowly helping many scientists by taking away the laborious task of manually annotating tissues, bones, cells and other structures. The work by Minnema *et al.* (4) shows an extreme case where manual segmentation is completely prohibitive for scalable scenarios.

Minnema *et al.* show the evaluation of a fully-automated CNN model developed for bone segmentation. The model was trained in a patch-wise manner using manually-annotated CT scans of patients who were treated with customized additive manufactured skull implants. Results of this evaluation show a Dice score of 0.92±0.04, suggesting that automatic and manual segmentations are very similar.

In comparison, a popular method for doing a semi-automated segmentation of CT scans is global thresholding, i.e., separate the bones and the background using an intensity threshold. Unfortunately, this approach is far from being perfect and often manual corrections by an expert are needed, which is time-consuming, user-biased and error-prone. Consequently, Minnema *et al.* concluded that the developed CNN model can offer an important opportunity to remove the prohibitive time and effort for bone segmentation in CT images, making additive manufacturing constructs affordable and accessible.

Deep learning also has some downsides though. One of the main barriers for the wide adoption of deep learning methods in clinical practice is usually caused by the variability in the data itself (e.g., contrast, resolution, signal-to-noise). Deep learning models usually suffer from a poor generalization when used with input data that comes from different machines (different vendor, model, etc.), with different acquisition parametrization or any underlying component that can cause the data distribution to shift. These over-parametrized models have a high tendency to rely on superficial statistical patterns of the data, and are not immune to the distributional shift caused by external factors (different scanner, different acquisition protocol, etc.).

This generalization gap can be mitigated through some techniques, which are not mutually-exclusive. One common denominator of all deep learning methods is that the more data there is to train a model, the better it will perform. While it is not easy for single sites to generate a large

© Journal of Medical Artificial Intelligence. All rights reserved.

amount of data and manual labels, multi-center initiatives and crowd sourcing are efficient approaches to federate such useful resources (5). Depending on the country, ethical and privacy limitations may occur. Another popular approach to solve the problem of generalization to other sites is to perform a domain adaptation (6), where data from another domain can be transferred to the current distribution thereby improving generalization. Other approaches rely on generative models, however, most prominent generative models such as Generative Adversarial Networks (GANs) (7) usually require a large amount of data and it can take shortcuts to model the underlying distribution. Recent likelihood-based models (8) showed some improvements, however, it is still very difficult to model such high-dimensional distributions. Another approach, also leveraging unlabeled data, is the use of semi-supervised (9) learning methods, that can yield improvements even in a small data regime.

From a different angle, one way to minimize the burden of creating manual annotations is to use active learning, whereby the most uncertain predictions are selected for manual correction before re-training the model (10).

While deep learning methods are encouraging, the outstanding issues noted above must be solved before they can be used confidently in clinical routine.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Journal of Medical Artificial Intelligence*. The article did not undergo external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2019.01.01). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
2. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60-88.
3. Chen LC, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv:1706.05587 [cs.CV] 2017.
4. Minnema J, van Eijnatten M, Kouw W, et al. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. Comput Biol Med 2018;103:130-9.
5. Zbontar J, Knoll F, Sriram A, et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. arXiv:1811.08839 [cs.CV] 2018.
6. Perone CS, Ballester P, Barros RC, et al. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. arXiv:1811.06042 [cs.CV] 2018.
7. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. arXiv:1406.2661 [stat.ML] 2014.
8. Menick J, Kalchbrenner N. Generating High Fidelity Images with Subscale Pixel Networks and Multidimensional Upscaling. arXiv:1812.01608 [cs.CV] 2018.
9. Perone CS, Cohen-Adad J. Deep semi-supervised segmentation with weight-averaged consistency targets. arXiv:1807.04657 [cs.CV] 2018.
10. Wang K, Zhang D, Li Y, Zhang R, Lin L. Cost-Effective Active Learning for Deep Image Classification. arXiv:1701.03551 [cs.CV] 2017.

doi: 10.21037/jmai.2019.01.01
Cite this article as: Perone CS, Cohen-Adad J. Promises and limitations of deep learning for medical image segmentation. J Med Artif Intell 2019;2:1.