Gastroenterology needs its own ImageNet

Fons van der Sommen

Eindhoven University of Technology, Eindhoven, The Netherlands

Correspondence to: Fons van der Sommen. Eindhoven University of Technology, Eindhoven, The Netherlands. Email: f.v.d.sommen@tue.nl.

Received: 28 October 2019; Accepted: 17 November 2019; Published: 13 December 2019.

doi: 10.21037/jmai.2019.11.03

View this article at: http://dx.doi.org/10.21037/jmai.2019.11.03

In a period spanning less than a decade, deep learning with Convolutional Neural Networks (CNNs) has become the standard for Artificial Intelligence (AI) applications. While most of the key concepts were introduced decades ago, the two driving forces behind its success have only started building momentum at the dawn of this century: (I) an exponential increase in computational power and (II) the growing availability of large sets of labeled data. These two developments reached critical mass first in speech recognition (1), and later in computer vision, with the introduction of AlexNet (2): the first deep learning architecture to win the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (3).

Ever since these first breakthroughs, the graphical processing units (GPUs) required for deep learning have been dropping in price and increasing in power, rendering the availability of data the limiting factor for deep learning. Especially in the medical domain, this is hampering the progress of AI, as medical data is hard to acquire and expensive to label. Moreover, additional challenges arise when building medical datasets, such as patient privacy, data uniformity over different medical centers and a considerable inter-observer variability among different medical experts, who provide the ground truth for training AI algorithms. Finally, data ownership is a rising topic of interest now that the first AI algorithms hit the market and start yielding commercial profit, while their development is completely fueled by patient data.

To deal with the scarcity of high-quality, labeled datasets, transfer learning has been first explored as an interesting solution for CNNs by Oquab *et al.* (4), and in 2016 some early successes in the medical domain were reported by Shin *et al.* (5) and Tajbakhsh *et al.* (6). This technique effectively alleviates the need for large, task-specific datasets to train deep neural networks by introducing a so-called pretraining stage, in which the network is trained on a large dataset from a different domain (for example natural images), for which an abundance of data is available. Then, during a second training stage, the target data (for example endoscopic images of neoplastic lesions) are used to fine-tune the model to the task at hand. In this fashion, the knowledge (i.e., network weights) that the network has learned from the large dataset is transferred to a different domain. The success behind this approach can be explained by the observation that a lot of visual features are useful for a variety of different image types. For example, visual features such as color/intensity transitions, edges, lines and basic shapes appear in both natural images (e.g., birds, cars and buildings) and in e.g., ophthalmology scans or stained histology slides. Moreover, the depth of the employed neural networks facilitates complex hierarchical patterns that are useful for a broad variety of tasks.

For the abovementioned reasons, nearly all methods proposed for AI-based medical image analysis exploit a form of transfer learning. To this end, ImageNet is by far the most popular dataset for pretraining purposes and it is used for a wide variety of tasks: from classification of pulmonary tuberculosis in chest radiography (7) to whole-slide pathology image analysis to detect of lymph node metastases in women with breast cancer (8). Also in endoscopy, ImageNet pretraining is used ubiquitously: for polyp localization (9) and staging (10), classification for invasion depth of esophageal squamous cell carcinoma (11) and the detection of early neoplasms in Barrett's esophagus (12-14).

While transfer learning using ImageNet as basis is widely employed, its efficiency can be questioned. The ImageNet dataset (15) consists of 1.2 million images of a thousand different categories, including "tree", "tool" and "tractor", but also more sophisticated categories such

as "membranophone", "face powder" and "mongoose"—not to mention the 120 dog breeds that have been added as categories for fine-grained classification[1]. Although ImageNet pretraining has proven to be an efficient tool for medical image analysis with AI, it is at least unsettling that a complete understanding of the inner workings driving this success is lacking. Moreover, disturbing results have been reported on fooling neural networks that were trained using ImageNet (16). Therefore, it is not surprising that ImageNet pretraining is questioned by He *et al.* in a recent publication (17) and there's an increasing interest in domain-specific pretraining (18,19). The reasoning behind this approach is twofold: (I) from a logical point of view, pretraining with images that are more similar to the target domain is more intuitive than pretraining with images of cats, dogs and trees, and, more importantly, (II) it leads to better and more robust results that are less prone to the natural variation exhibited by the images in the target domain (e.g., endoscopic imagery).

In an attempt to demonstrate the effectiveness of domain-specific pretraining in endoscopy, our consortium[2] has constructed a set of 494,355 retrospectively collected endoscopic images and have roughly categorized a subset of 3,743 images into five classes (i.e., "stomach", "duodenum", "esophagus", "colon" and "other"). Our first experiments indicate that even with a small proportion of the images labeled and a very modest number of categories, an improvement is shown for the detection of early Barrett's dysplasia (20). Moreover, further experiments demonstrated that this dataset also offers a suitable basis for an AI algorithm to grade the informativeness of endoscopic video frames (21). We hypothesize that this dataset, which we have titled GastroNet, serves as a better pretraining dataset for endoscopic imaging problems than the widely employed ImageNet and we are currently setting up the experiments that aim to test this hypothesis.

From our experiments, a number of first observations can be made: (I) domain-specific pretraining can lead to improved results of AI algorithms within endoscopy, (II) the power harnessed by such a domain-specific dataset is not only limited by the number of labeled samples, but also by the number of classes, as this forces the neural network to forge complex hierarchical structures with high discriminative power and (III) partially labeled, poorly structured, heterogenous data can be beneficial for pretraining deep neural networks. Especially the latter observation is interesting, since most medical data satisfies these conditions. However, dealing with this data to unlock its potential comes with substantial technical challenges and introduces numerous research questions regarding the design of deep neural networks and how they are trained. How to deal with missing labels, exploit variability in the data and combine different types of data during the training phase, just to name a few.

Over the past 5 years, AI-assisted endoscopy has grown from a curiosity into a valuable asset that will shape the future of gastroenterology. About a decade ago, the first outlines of this field became slowly visible by retrospective experiments on small, homogeneous datasets (22-24), recently, large clinical, prospective studies on real-time endoscopic video demonstrate that the field has matured (25,26). To maintain this momentum, and to make the next step into clinical application, additional parameters become important, such as, for example, robustness against small variations over different imaging devices and interpretability of the AI predictions. In this next development stage, data will play an increasingly important role, for both training algorithms and validating their clinical performance. Given the costly acquisition of high-quality labeled datasets, we anticipate that retrospectively collected, largely unlabeled data will fuel further progress of the field. In particular, the abundance of unlabeled endoscopic image data will facilitate domain specific pre-training, leading to a domain-native neural network that is more robust against the natural variations within endoscopic imagery.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the Guest Editors (Michael F. Byrne and Brandon J. Teng) for the series "Artificial Intelligence and Gastrointestinal Cancer" published in *Journal of Medical*

[1] A very interesting video can be found online about what ImageNet pretrained neural networks "see" when the neuron excitations are amplified during a clip of a grocery trip, and it's a lot of animals and in particular—not surprisingly—dogs.

[2] The University Medical Centers (UMC) Amsterdam, the Catharina Hospital Eindhoven and Eindhoven University of Technology.

Artificial Intelligence. The article did not undergo external peer review.

Conflicts of Interest: The author has completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2019.11.03). The series "Artificial Intelligence and Gastrointestinal Cancer Column" was commissioned by the editorial office without any funding or sponsorship. The author has no other conflicts of interest to declare.

Ethical Statement: The author is 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.

References

1. Dahl GE, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Process 2012;20:30-42.
2. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012.
3. Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 2015;115:211-52.
4. Oquab M, Bottou L, Laptev I, et al. Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition; 23-28 June 2014; Columbus, OH, USA; IEEE, 2014:171724.
5. Shin HC, Roth HR, Gao M, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging 2016;35:1285-98.
6. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? IEEE Trans Med Imaging 2016;35:1299-312.
7. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology 2017;284:574-82.
8. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017;318:2199-210.
9. Urban G, Tripathi P, Alkayali T, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology 2018;155:1069-1078.e8.
10. Fonollá R, van der Sommen F, Schreuder RM, et al. Multi-Modal Classification of Polyp Malignancy using CNN Features with Balanced Class Augmentation. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019); 8-11 April 2019; Venice, Italy, Italy; IEEE, 2019:74-8.
11. Nakagawa K, Ishihara R, Aoyama K, et al. Classification for invasion depth of esophageal squamous cell carcinoma using a deep neural network compared with experienced endoscopists. Gastrointest Endosc 2019;90:407-14.
12. Van Riel S, Van Der Sommen F, Zinger S, et al. Automatic detection of early esophageal cancer with CNNS using transfer learning. 2018 25th IEEE International Conference on Image Processing (ICIP); 7-10 Oct. 2018; Athens, Greece; IEEE, 2018:1383-7.
13. Ghatwary N, Ye X, Zolgharni M. Esophageal Abnormality Detection Using DenseNet Based Faster R-CNN with Gabor Features. IEEE Access 2019;7:84374-85.
14. van der Putten J, Wildeboer R, de Groof J, et al. Deep learning biopsy marking of early neoplasia in Barrett's esophagus by combining WLE and BLI modalities. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019); 8-11 April 2019; Venice, Italy, Italy; IEEE, 2019:1127-31.
15. Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition 2009:248-55. doi:10.1109/CVPRW.2009.5206848.
16. Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. arXiv:1312.6199 [cs.CV]. 2013.
17. He K, Girshick R, Dollár P. Rethinking imagenet pre-training. arXiv:1811.08883 [cs.CV]. 2018.

18. Cui Y, Song Y, Sun C, et al. Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:4109-18. doi:10.1109/CVPR.2018.00432
19. Tschandl P, Sinz C, Kittler H. Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation. Comput Biol Med 2019;104:111-6.
20. van der Putten J, de Groof J, van der Sommen F, et al. Pseudo-labeled Bootstrapping and Multi-stage Transfer Learning for the Classification and Localization of Dysplasia in Barrett's Esophagus. In: Suk HI, Liu M, Yan P, et al. editors. Machine Learning in Medical Imaging. MLMI 2019. Lecture Notes in Computer Science, vol 11861. Springer, Cham, 2019:169-77.
21. van der Putten J, de Groof J, van der Sommen F, et al. Informative Frame Classification of Endoscopic Videos Using Convolutional Neural Networks and Hidden Markov Models. 2019 IEEE International Conference on Image Processing (ICIP); 22-25 Sept. 2019; Taipei, Taiwan, Taiwan; IEEE, 2019:380-4.
22. Maroulis DE, Iakovidis DK, Karkanis SA, et al. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. Comput Methods Programs Biomed 2003;70:151-66.
23. Kodogiannis VS, Boulougoura MG, Lygouras J, et al. A Neuro-Fuzzy-Based System for Detecting Abnormal Patterns in Wireless-Capsule Endoscopic Images. Neurocomputing 2007;70:704-17.
24. van der Sommen F, Zinger S, Schoon EJ, et al. Computer-Aided Detection of Early Cancer in the Esophagus using HD Endoscopy Images. Proceedings Volume 8670, Medical Imaging 2013: Computer-Aided Diagnosis; 86700V (2013).
25. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019;68:94-100.
26. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813-9.

doi: 10.21037/jmai.2019.11.03
Cite this article as: van der Sommen F. Gastroenterology needs its own ImageNet. J Med Artif Intell 2019;2:23.