Guest Editorial Annotation-Efficient Deep Learning: The Holy Grail of Medical Imaging

I. Introduction

ANNOTATION-EFFICIENT deep learning refers to methods and practices that yield high-performance deep learning models without the use of massive carefully labeled training datasets. This paradigm has recently attracted attention from the medical imaging research community because (1) it is difficult to collect large, representative medical imaging datasets given the diversity of imaging protocols, imaging devices, and patient populations, (2) it is expensive to acquire accurate annotations from medical experts even for moderately sized medical imaging datasets, and (3) it is infeasible to adapt data-hungry deep learning models to detect and diagnose rare diseases whose low prevalence hinders data collection.

The challenge of annotation-efficient deep learning has been approached from various angles in the medical imaging literature; however, the relevant publications are scattered across numerous sources and there exist many gaps that require further research. This issue addresses these deficiencies by presenting a collection of state-of-the-art research articles spanning the major topics of annotation-efficient deep learning, which are diagrammed in Figure 1.

The call for papers attracted significant interest from the medical imaging community. A total of 101 manuscripts were submitted and 32 were finally accepted for publication. The submissions were subjected to a rigorous review process, in which each manuscript was refereed by 3 to 6 experts in the field and underwent typically two rounds of revision. Figure 2 graphs the number of accepted manuscripts related to each topic and its subtopics. Most of the articles fall in the categories of leveraging unannotated data and utilizing annotations efficiently. The most popular subtopics include zero/few-shot learning, domain adaptation, learning from weak and noisy labels, and synthetic data augmentation. In the remainder of this editorial, we will first categorize and summarize the articles and then highlight potential opportunities for future research. For this editorial to be self-contained, we list all articles included in this Special Issue in the appendix and refer to each of them with "A" followed by its item number.

II. Acquiring Annotations Efficiently

Active learning and interactive segmentation are two common practices for acquiring annotations in a budget-efficient manner. The former determines which data samples should be annotated, while the latter shortens the annotation session. The two approaches are complementary, and they enable substantial reductions in annotation time and cost.

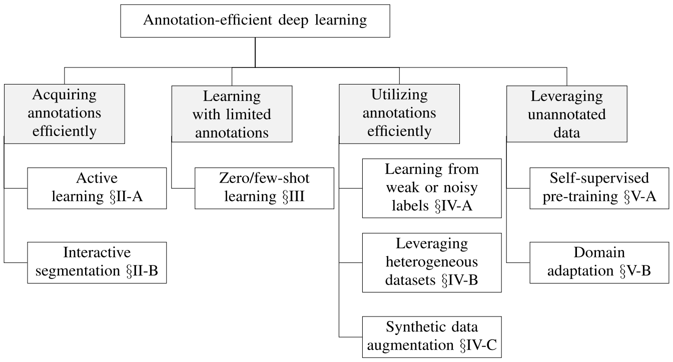
Annotation-efficient deep learning
Acquiring annotations efficiently
Learning with limited annotations
Utilizing annotations efficiently
Leveraging unannotated data
Active learning §II-A
Interactive segmentation §II-B
Zero/few-shot learning §III
Learning from weak or noisy labels §IV-A
Leveraging heterogeneous datasets §IV-B
Synthetic data augmentation §IV-C
Self-supervised pre-training §V-A
Domain adaptation §V-B

Fig. 1. Overview of the topics covered in this Special Issue on annotation-efficient deep learning.

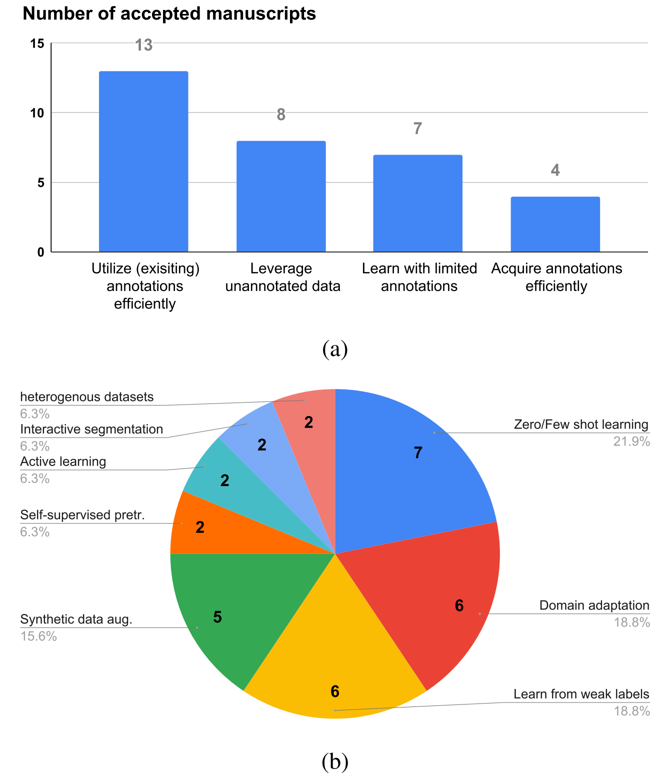

Fig. 2. Number of articles related to each of the (a) four major topics and (b) all subtopics.

This work was supported in part by the National Institutes of Health (NIH) under Award R01HL128785.
Digital Object Identifier 10.1109/TMI.2021.3089292

A. Active Learning

Active learning aims to select the most informative and representative samples for experts to annotate, thereby minimizing the total number of samples that must be annotated to train performant models. The effectiveness of an active learning method largely depends on its criteria for determining the informativeness and representativeness of an unlabeled sample. The more uncertain the sample's prediction, the more information its label offers. Minimizing the number of annotated samples requires that the labeled samples be distinct from one another. Therefore, uncertainty and diversity are two natural metrics for informativeness and representativeness [1], [2], upon which the two articles featured in this Special Issue present two methods. Nath *et al.* [A1] argue that duplicating uncertain samples in the labeled training dataset helps decrease the overall model uncertainty, and they regulate the amount of duplication based on mutual information between data pairs of the training and unlabeled pool. Mahapatra *et al.* [A2] propose a self-supervised method for training a classifier to select informative samples based on saliency maps that embody both uncertainty and diversity.

B. Interactive Segmentation

Smart interactive segmentation tools play an essential role in reducing the manual burden of producing high-quality annotated data. In this Special Issue, two teams have made advances in this topic. Ma *et al.* [A3] explore the use of boundary-aware multiagent reinforcement learning to interactively and iteratively refine 3-D image segmentations. The approach can combine different types of user interactions, such as point clicks and scribbles, via an efficient "supervoxel-clicking" design. Feng *et al.* [A4] investigate few-shot learning to arrive at better medical image segmentation using only limited supervision. The approach has the potential to reduce the annotation burden and fix some common issues in few-shot learning methods by enabling the user to make minor corrections interactively.

III. Learning With Limited Annotations

The ability to analyze medical images with little or no training data is highly desirable. For example, Huang *et al.* [A5] present an unsupervised deep learning method for multicontrast MR image registration, using a coarse-to-fine network architecture consisting of affine coarse and deformable fine transformations to achieve end-to-end registration. Huang *et al.* [A6] suggest an image denosing method based on unsupervised learning, which disentangles image and noise content without requiring paired noisy and normal images. For the task of semantic segmentation, compiling large-scale (nonsynthetic) medical image datasets with pixel-level annotations is time-consuming and often prohibitively expensive, and it may be impossible to balance all the relevant classes in the training set. Although semi-supervised approaches aim to relax the level of supervision to bounding boxes and image-level tags, these models still require copious training samples and are prone to suboptimal performance on unseen classes.

A. Zero/Few-Shot Learning

By contrast, the few-shot learning paradigm attempts to utilize a few annotated "support" samples to learn representations of unseen classes, denoted "query" samples. The few-shot learning paradigm was initially focused on classification and later applied to segmentation. In the context of volumetric ultrasound segmentation, Al Chanti *et al.* [A7] introduce a decremental update of the objective function to guide model convergence in the absence of copious annotated data. Their update strategy switches from weakly supervised training to a few-shot setting. Also in the context of volumetric segmentation, in this case of the heart, Wang *et al.* [A8] propose a few-shot learning framework that combines ideas from semi-supervised learning and self-training. The key to their success lies in cascaded learning; specifically, in the selection and evolution of high-quality pseudo labels produced by an auto-encoder network that learns shape priors. Feng *et al.* [A4] introduce interactive learning into the few-shot learning strategy for image segmentation. In their approach, an expert indicates error regions as an error mask that corrects the predicted segmentation mask from which the network learns to produce better segmentations. Paul *et al.* [A9] propose a zero-shot learning strategy for the diagnosis of chest radiographs, leveraging multiview semantic embedding and incorporating self-training to deal with noisy labels. Cui *et al.* [A10] propose a unified framework for generalized low-shot medical image segmentation that deals both with data scarcity and lack of annotations, which is the case for rare diseases. Using distance metric learning, their model learns a multimodal mixture representation for each category and effectively avoids overfitting the extremely limited data.

Finally, a recent trend in image segmentation is combining traditional model-based contour segmentation methods with deep learning methods to accrue the benefits of both approaches [3]. Accordingly, Lu *et al.* [A11] formulate anatomy segmentation as a contour evolution process, using Active Contour Models whose evolution is governed by graph convolutional networks. The model requires only one labeled image exemplar and supports human-in-the loop editing. Their one-shot learning scheme leverages additional unlabeled data through loss functions that measure the global shape and appearance consistency of the contours.

IV. Utilizing Annotations Efficiently

Acquiring additional strongly annotated data is arguably the best approach to improving deep learning models, but this practice may not always be feasible due to limited budgets or shortages of medical expertise. Consequently, one may resort to using annotations that are cheaper or faster to obtain. The resulting weak and noisy annotations require proper learning methodologies (Section IV-A). Another approach to this challenge is to leverage related annotated datasets, which can increase the effective size of training sets (Section IV-B). Synthetic data augmentation (Section IV-C) is yet another popular approach by which the training sets are amplified by creating artificial examples and corresponding annotations.

A. Learning From Weak or Noisy Labels

Enabling learning from weak or noisy labels can be especially useful for medical imaging as the collection of quality labels is time-consuming, cumbersome, and often requires expert knowledge. This was one of the most popular topics of this Special Issue, probably due to the availability of weak or noisy labels that are routinely recorded during clinical workflows, such as measurements taken by clinicians (e.g., RECIST) or image-level labels provided in clinical reporting. Authors have addressed the topic by exploiting both "weak" labels (e.g., image-level labels [4] instead of pixel/voxel-wise labels) or "noisy" labels transferred from other modalities, and demonstrated promise by effectively combining supervision from such weak or noisy signals to train medical imaging models.

The research direction most investigated in this Special Issue is the utilization of weak image-level labels. For example, Liu *et al.* [A12] employ a self-supervised attention mechanism to learn from image-level labels for pediatric bone age assessment. Ouyang *et al.* [A13] exploit weak supervision from image-level labels to train abnormality localization models for chest X-ray diagnosis. Tardy and Mateus [A14] use image-level labels from mammography to detect abnormalities in a weakly self-supervised fashion. Wang *et al.* [A15] combine weakly supervised and semi-supervised learning to train models from both image-level and densely labeled images. Both attention guidance and multiple-instance learning are utilized to effectively learn from both types of annotated data for the purposes of adenocarcinoma prediction in abdominal CT. Another interesting research direction is presented by Zhao and Yin [A16], who explore the use of point annotations for weakly supervised cell segmentation in microscopy images. Using one point per cell, they train segmentation models that perform comparably to models trained on dense annotation maps. The utilization of noisy labels is explored by Ding *et al.* [A17], who use weak-supervision for automated vessel detection in ultra wide-field fundus images. They exploit a deep neural network pretrained on a different imaging modality, fluorescein angiography, together with multimodel registration to iteratively train their model and simultaneously refine the registration.

B. Leveraging Heterogeneous Datasets

Another promising approach to mitigating annotation costs in medical imaging is to leverage multiple heterogeneous datasets. These datasets might have been acquired for other purposes, but they can be combined to build more robust models. Two articles explore this approach. Yan *et al.* [A18] train a lesion detection ensemble from multiple heterogeneous lesion datasets in a multitask approach (their test dataset, including manually annotated 3-D lesions in CT, has been publicly released). The ensemble is utilized for proposal fusion and missing annotations are mined from partially labeled datasets to improve the overall detection sensitivity. Li *et al.* [A19] exploit intra-domain and inter-domain knowledge to improve cardiac segmentation models. They evaluate their method on a multimodality dataset, demonstrating improvements over previous semi-supervised and domain adaptation methods. Both articles are promising with regard to the future development of AI models because it is natural to combine information about human anatomy acquired for different purposes. These approaches may be able to leverage features across datasets in order to build more powerful and robust models, as well as to reduce drastically the cost of curating task-specific datasets.

C. Synthetic Data Augmentation

Synthetic data augmentation aims to generate artificial yet realistic-looking images, thereby enabling the construction of large and diverse datasets. This practice is particularly desirable for organ segmentation where acquiring large labeled datasets is expensive and time-consuming and, even more importantly, for lesion segmentation where data collection is hindered by the low prevalences of underlying diseases and conditions. This Special Issue features five articles demonstrating that the addition of synthetic data to real data either improves segmentation performance or enables a comparable level of performance while reducing the need for real annotated data. These articles cover various medical imaging modalities, including MR scans, CT scans, ultrasound images, microscopic images, and retinal images. The proposed methods differ in the training data required to synthesize images and their corresponding segmentation masks. Relying on variants of the CycleGAN, three articles exploit image synthesis for the task of image segmentation but take different measures to circumvent the use of segmentation masks from the target domain. Specifically, Gilbert *et al.* [A20] use a preexisting 3-D shape model of the heart, Wang *et al.* [A21] require segmentation masks from a domain only related to the target domain, and Yao *et al.* [A22] rely on a handcrafted model to generate synthetic COVID lesions. The other two articles do require segmentation masks from the desired domain in order to generate synthetic data. In particular, Liu *et al.* [A23] generate synthetic segmentation masks through an active appearance model, which is trained using segmentation masks from the target domain, and Guo *et al.* [A24] rely on brain atlases, which makes their method suitable for applications where high-quality atlases are available for training. Despite the promising performances of the aforementioned methods, the requirement for atlases or shape and appearance models of the target diseases or organs may limit the applicability.

V. Leveraging Unannotated Data

Unannotated images contain a wealth of knowledge that can be leveraged in various settings, such as self-supervised learning (Section V-A) and unsupervised domain adaptation (Section V-B). The former utilizes unannotated data to pretrain the model weights, whereas the latter leverages unannotated data from a target domain to mitigate the distribution shifts between the training and test datasets.

A. Self-Supervised Learning

Self-supervised learning has recently gained prominence for its capacity to learn generalizable and transferable representations without the use of any expert annotation. The idea is to pretrain models on pretext tasks (e.g., rotation [5], inpainting [6], and contrasting [7]), where supervisory signals are automatically derived directly from the unlabeled data themselves (avoiding expert annotation costs), and then fine-tune (transfer) the pretrained models in a supervised manner so as to achieve annotation efficiency for the target tasks (e.g., segmenting organs and localizing diseases). Self-supervised learning often leverages the underlying structures and intrinsic properties of the unlabeled data, a feature that is particularly attractive in medical imaging. To take advantage of the special property of pathology, Koohbanani *et al.* [A25] design pathology-specific pretext tasks and demonstrate annotation efficiency for the target domain in the small data regime. To exploit the semantics of anatomical patterns embedded in medical images, Haghighi *et al.* [A26] pretrain their model using chest CT and demonstrate its cross-domain capability to contrast-enhanced CT and MRI. They further show that, as an add-on strategy, their method complements existing self-supervised methods (e.g., inpainting [6], context restoration [8], rotation [5], and models genesis [9], [10]), thus boosting performance.

B. Domain Adaptation

Deep learning models struggle to generalize when the target domain exhibits a data distribution shift with respect to the source dataset used for training. This challenge is even more pronounced in the field of medical imaging where variations in ethnicities, scanner devices, and imaging protocols lead to large data distribution shifts. Unsupervised domain adaption has emerged as an effective approach to improving the tolerance of deep learning models to the distribution shifts in medical imaging datasets. The Special Issue features six articles on this topic. Mobiny *et al.* [A27] tackle the distribution shift through an episodic training strategy where the training data of each episode are generated so as to mimic a distribution shift with respect to the original training dataset. Xing *et al.* [A28] enhance detection performance on the target dataset using a bidirectional GAN and pseudo-generated labels. The other four articles improve the CycleGAN in domain adaptation. Tomczak *et al.* [A29] include auxiliary segmentation and reconstruction tasks. Ju *et al.* [A30] exercise consistency regularization during training. Tomar *et al.* [A31] introduce a self-attentive spatial normalization block in the generator networks. Koehler *et al.* [A32] use task-specific probabilities to focus the attention of the transformation module on more relevant regions. In these articles, domain adaptation is leveraged for various applications, including reorienting MR images, staining unstained white blood cell images, cross-modality heart chamber segmentation between unpaired MR and CT scans, nucleus detection in cross-modality microscopy images, lung nodule detection in cross-protocol CT datasets, and various medical vision applications in cross-domain retinal imaging.

VI. Research Opportunities

It was inspiring to see the numerous submissions to this Special Issue and the quality of the accepted manuscripts. However, there remain unanswered questions and unaddressed issues that offer many enticing research opportunities.

A. Quantifying Annotation Efficiency

Annotations generally refer to the ground truth information that is used in training, validating, and testing models. In medical imaging, this information is mostly provided (subjectively) by experts, but it may also derive from objective conditions obtained through tests (e.g., the malignancy of a tumor) and from medical concepts (diseases and conditions) automatically extracted from clinical notes and diagnostic reports. In a broader sense in self-supervised learning, it could also be a part of the data that are to be predicted based on other parts of the data or the original data that are to be restored from their transformed versions [10]. Annotations can be acquired at the patient, image, and pixel levels. Therefore, they require different levels of effort/cost and offer different levels of semantics/power as supervisory signals in training models. With regard to annotation, Method A is said to be more efficient than Method B if, compared with Method B, Method A (1) achieves better performance with equal annotation cost, (2) offers equal performance with lower annotation cost, or (3) provides equal performance with equal annotation cost, but reduces the training time. Currently, the prevailing literature generally assumes the same annotation cost for each and every sample (e.g., patient, image, or lesion), but costs could vary dramatically from one sample to another. It is also important to understand the trade-off between annotation cost and supervisory power in different settings. For example, for proton therapy, a few lesions with carefully delineated masks may have more supervisory power than many lesions coarsely bounded by boxes. There is a need for new concepts, algorithms, and tools to analyze annotation efficiency across different contexts.

B. Annotating Patients Efficiently

This matter is essential to medical image analysis, and resolving it can greatly accelerate the development of deep learning models. The Special Issue offers only four articles on this topic, leaving it under-investigated. In particular, the two articles on active learning primarily focus on the tasks of image segmentation and classification, leaving other important tasks such as lesion detection largely unexplored. This application bias is not peculiar; in fact, it applies to the entire medical imaging literature as well. It could be due to the fact that finding an optimal set of samples that are informative and representative is inherently difficult [11], [12]. With regard to interactive segmentation, the issue features two articles with promising task-specific solutions. However, generic interactive segmentation tools remain difficult to build for medical imaging applications. The type of user interactions employed (points, scribbles, bounding boxes, polygons, etc.) is often based on the targeted anatomy, which tends to hinder generalization to new anatomical structures. Building more comprehensive datasets with which to train general-purpose interactive models on several target anatomies and/or imaging modalities may be a good way forward. It would be desirable to develop tools that integrate active learning with interactive segmentation, as active learning can select the most important samples for annotation while interactive segmentation can shorten the annotation session for each sample. Active learning also may be embedded within interactive segmentation to suggest which parts of the image should be segmented next, thereby further accelerating the annotation process. Such tools will prove to be not only valuable for clinical purposes but also indispensable for building massive, strongly annotated datasets—essential infrastructure for research, without which the annotation efficiency of a method can hardly be quantitatively determined or benchmarked.

C. Learning by Zero/Few Shots

A majority of the articles contained in this Special Issue focus on zero- or few-shot learning for medical image classification and segmentation, reflecting the eagerness in the community to progress beyond data-intensive machine learning methods. The promise of this technology is especially appealing in the healthcare domain where the collection of large and varied annotated datasets is difficult and sometimes hindered by regulations, a problem that is aggravated when studying rare diseases for which the acquisition of training examples becomes even more difficult. In future research, it may be beneficial to consider the use of multimodality information in designing few-shot learning methods. One could force the embeddings learned from multiple modalities (e.g., X-rays, CT, MRI, and reports) to be matched in a shared space, thereby encouraging collaborative learning via joint- and cross-supervision. Furthermore, analogous to the saying that training to identify counterfeit currency begins with studying genuine money, an effective approach to zero- or few-shot learning for diagnosing diseases and abnormal conditions in medical imaging could begin with learning dense (normal) anatomical embeddings. Successful zero- and few-shot learning will bring us closer to human abilities, i.e., a clinician learning to identify/diagnose certain diseases by studying just a few textbook examples.

D. Synthesizing Annotations

Artificially generating realistic-looking images with associated ground truth information helps relieve laborious human annotation and facilitates the creation of large datasets. This is particularly attractive for image segmentation where acquiring carefully delineated masks is tedious and time-consuming as well as for localization, where collecting rare diseases and conditions is challenging. Hand-crafted or trained generative models have proven to be powerful in creating “realistic” images and videos. Nevertheless, for the purposes of medical imaging, special attention and care must be given to potential artifacts. Embedding physiological and anatomical knowledge as well as the physical principles of imaging modalities into the synthesis process may prove to be critical.

E. Innovating Architectures

Recent advances in network architecture design and search have been proved successful in natural language processing and computer vision. However, it is not clear whether these new architectures are more annotation-efficient than their predecessors. For instance, Tan and Le [13] show that transformer architectures [14] as well as a new family of efficient models, called EfficientNet v2.0, benefit from a significantly larger version of ImageNet, which raises concerns about the annotation efficiency of new architectures. Further research is required to study the annotation efficiency of such models, particularly in the context of medical imaging. It would be intriguing to investigate if, by reducing the need for annotated data, annotation-efficient methods can engender new architectural advancements. For instance, Caron *et al.* [15] demonstrate that self-supervised training is effective in reducing the amount of annotated visual data required to train transformers. It is worth studying how such recent architectural advancements benefit from other annotation-efficient paradigms in the context of medical imaging. Although the medical imaging community tends to adopt deep architectures developed for computer vision, given the differences between medical and natural images, to maximize annotation utilization [16]–[18], it would be worth designing or automatically searching for network architectures that exploit the particular opportunities of medical imaging for annotation efficiency [19]. It would be fascinating to explore and develop new computing architectures and paradigms (e.g., quantum computing) for annotation efficiency in medical imaging.

F. Exploiting Clinical Information

Medical concepts (diseases and conditions) extractable from clinical notes and reports may be capitalized as (weak) patient-level annotation [4]. Indeed, natural language processing (NLP) has already been utilized to harvest image-level annotations from diagnostic reports in order to build medical imaging datasets and used to develop machine learning models [20], but the vast resources of clinical notes and diagnostic reports that are available in electronic health records of hospital systems have yet to be fully mined and well utilized. Learning representations jointly from both images and texts (clinical notes and diagnostic reports) will likely be a very active area of research in the near future, although patient privacy and other regulatory constraints must be carefully considered [21].

G. Integrating Data and Annotation From Multiple Sources

Datasets created at different institutions tend to be annotated differently even when addressing the same clinical issue. There remains a need for learning methods that can seamlessly integrate data and annotation from different sources. Federated learning (FL) has recently emerged as an effective solution for addressing the growing concerns about data privacy when integrating data and annotation from different providers. FL trains a model using data from various sites without breaching patient privacy and other regulations [22], thereby making more data available for AI model development. FL has yet to seamlessly handle the unavoidable heterogeneity of data across different sites. Semi-supervised [23] and unsupervised/self-supervised [24] approaches have already been successfully used in the context of FL. Nevertheless, it will be interesting to see how other annotation-efficient methods can be combined with FL as well as how FL and other methods that integrate data and annotation from multiple sources can relieve the annotation burden.

H. Mining Common Knowledge

Currently, the dominant approach in deep learning—supervised learning—offers expert-level and sometimes even super-expert-level performance. Models trained via supervised learning have also demonstrated remarkable capacity in knowledge transfer across domains [25], [26], but at their core, they are trained to be "specialists" [27] on (target) tasks that can be annotated by experts. There are many diseases and conditions in medical imaging as well as common medical knowledge that can hardly be annotated even by willing experts. Self-supervised learning has proven to be promising in training models to be "generalists" on *various* pretext tasks so as to reduce expert annotation effort on target tasks. Typically, the semantics of expert-provided annotation is strong but narrow (task-specific), while that of machine-generated annotation in self-supervised learning is weak but general. Fundamental to annotation efficiency is learning common generalizable knowledge. Therefore, self-supervised learning will inevitably overtake supervised learning in extracting generalizable and transferable knowledge; i.e., self-supervised representation learning followed by (supervised) transfer learning is poised to become the most practical paradigm toward annotation efficiency. This is particularly true for medical imaging, because medical images harbor rich semantics about human anatomy, thus offering a unique opportunity for deep semantic representation learning. Yet, harnessing the powerful semantics associated with medical images remains largely unexplored [28], [29]. Furthermore, medical images are often augmented by clinical notes and reports, making it even more attractive to learn generic semantic representations jointly from both images and reports via self-supervision. Minding common knowledge will prove to be impactful on learning by zero/few shots, accelerating model training (supervised, self-supervised, semi-supervised, unsupervised, federated, etc.), creating dense (normal) anatomical embeddings, characterizing diseases, their sub-types, and their semblances (false positives), and enhancing system performance and robustness.

I. Reusing Knowledge in Trained Models

Research and development in deep learning across academia and industry have resulted in numerous models trained on various datasets in supervised, self-supervised, semi-supervised, unsupervised, or federated manners for diverse clinical objectives. These trained models retain a large body of knowledge, and properly reusing this knowledge could reduce annotation efforts and accelerate training cycles, thereby increasing annotation efficiency. However, current practice in reusing knowledge from (existing) pretrained models for new tasks is very limited. Therefore, advanced methods are needed for transferring, reusing, and distilling the knowledge [30] from pretrained models as well as integrating the knowledge from multiple pretrained models with the same or distinct architectures.

J. Demonstrating Annotation Efficiency in Practice

Methods and techniques are being developed from various perspectives to circumvent the annotation dearth in medical imaging; their value and effectiveness need to be quantitatively evaluated and benchmarked. This calls for an infrastructure of massive, strongly annotated datasets in well-defined domains (e.g., pulmonary embolism [31], [32], colon cancer [33], [34], and cardiovascular disease [35], [36]), without which the annotation efficiency of a method cannot be adequately understood. As a result, we expect large-scale competitions and challenges to be organized. It would be encouraging to see more reports that annotation-efficient methods and practices have been employed in the design and development of commercial products.

There is no doubt that deep learning has dramatically transformed medical imaging; annotation-efficient deep learning remains the Holy Grail of medical imaging.

Acknowledgments

We would like to thank all the authors who submitted manuscripts to this Special Issue. We greatly appreciate the valuable efforts of the many expert reviewers who contributed to the selection of articles that have shaped the issue. We would also like to thank Fatemeh Haghighi, Jeffrey Liang, Julia Liang, and Zongwei Zhou for their valuable comments on earlier drafts. We are grateful to the Editor-in-Chief, Prof. Leslie Ying, for approving our proposal to edit this special issue and for her guidance along the way. Our special thanks go to the Managing Editor, Prof. Rutao Yao, for his timely and helpful support.

Nima Tajbakhsh, *Guest Editor*
VoxelCloud, Inc.
Los Angeles, CA 90024 USA

Holger Roth, *Guest Editor*
NVIDIA, Inc.
Bethesda, MD 20814 USA

Demetri Terzopoulos, *Guest Editor*
Computer Science Department
University of California at Los Angeles
Los Angeles, CA 90095 USA
VoxelCloud, Inc.
Los Angeles, CA 90024 USA

Jianming Liang, *Guest Editor*
Biomedical Informatics Program
Arizona State University
Tempe, AZ 85281 USA

Appendix
Related Works

[A1] V. Nath, D. Yang, B. A. Landman, D. Xu, and H. R. Roth, "Diminishing uncertainty within the training pool: Active learning for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2534–2547, Oct. 2021.

[A2] D. Mahapatra, A. Poellinger, L. Shao, and M. Reyes, "Interpretability-driven sample selection using self supervised learning for disease classification and segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2548–2562, Oct. 2021.

[A3] C. Ma *et al.*, "Boundary-aware supervoxel-level iteratively refined interactive 3D image segmentation with multi-agent reinforcement learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2563–2574, Oct. 2021.

[A4] R. Feng *et al.*, "Interactive few-shot learning: Limited supervision, better medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2575–2588, Oct. 2021.

[A5] W. Huang *et al.*, "A coarse-to-fine deformable transformation framework for unsupervised multi-contrast MR image registration with dual consistency constraint," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2589–2599, Oct. 2021.

[A6] Y. Huang *et al.*, "Noise-powered disentangled representation for unsupervised speckle reduction of optical coherence tomography images," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2600–2614, Oct. 2021.

[A7] D. A. Chanti, V. G. Duque, M. Crouzier, A. Nordez, L. Lacourpaille, and D. Mateus, "IFSS-net: Interactive few-shot Siamese network for faster muscle segmentation and propagation in volumetric ultra sound," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2615–2628, Oct. 2021.

[A8] W. Wang *et al.*, "Few-shot learning by a cascaded framework with shape-constrained pseudo label assessment for whole heart segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2629–2641, Oct. 2021.

[A9] A. Paul *et al.*, "Generalized zero-shot chest X-ray diagnosis through trait-guided multi-view semantic embedding with self-training," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2642–2655, Oct. 2021.

[A10] H. Cui, D. Wei, K. Ma, S. Gu, and Y. Zheng, "A unified framework for generalized low-shot medical image segmentation with scarce data," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2656–2671, Oct. 2021.

[A11] Y. Lu *et al.*, "Contour transformer network for one-shot segmentation of anatomical structures," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2672–2684, Oct. 2021.

[A12] C. Liu, H. Xie, and Y. Zhang, "Self-supervised attention mechanism for pediatric bone age assessment with efficient weak annotation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2685–2697, Oct. 2021.

[A13] X. Ouyang *et al.*, "Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2698–2710, Oct. 2021.

[A14] M. Tardy and D. Mateus, "Looking for abnormalities in mammograms with self-and weakly supervised reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2711–2722, Oct. 2021.

[A15] Y. Wang, P. Tang, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille, "Learning inductive attention guidance for partially supervised pancreatic ductal adenocarcinoma prediction," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2723–2735, Oct. 2021.

[A16] T. Zhao and Z. Yin, "Weakly supervised cell segmentation by point annotation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2736–2747, Oct. 2021.

[A17] L. Ding, A. E. Kuriyan, R. S. Ramchandran, C. C. Wykoff, and G. Sharma, "Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2748–2758, Oct. 2021.

[A18] K. Yan *et al.*, "Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2759–2770, Oct. 2021.

[A19] K. Li, S. Wang, L. Yu, and P.-A. Heng, "Dual-teacher++: Exploiting intra-domain and inter-domain knowledge with reliable transfer for cardiac segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2771–2782, Oct. 2021.

[A20] A. Gilbert, M. Marciniak, C. Rodero, P. Lamata, E. Samset, and K. McLeod, "Generating synthetic labeled data from existing anatomical models: An example with echocardiography segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2783–2794, Oct. 2021.

[A21] L. Wang, D. Guo, G. Wang, and S. Zhang, "Annotation-efficient learning for medical image segmentation based on noisy pseudo labels and adversarial learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2795–2807, Oct. 2021.

[A22] Q. Yao, L. Xiao, P. Liu, and S. K. Zhou, "Label-free segmentation of COVID-19 lesions in lung CT," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2808–2819, Oct. 2021.

[A23] J. Liu *et al.*, "Active cell appearance model induced generative adversarial networks for annotation-efficient cell segmentation and identification on adaptive optics retinal images," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2820–2831, Oct. 2021.

[A24] P. Guo, P. Wang, R. Yasarla, J. Zhou, V. M. Patel, and S. Jiang, "Anatomic and molecular MR image synthesis using confidence guided CNNs," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2832–2844, Oct. 2021.

[A25] N. A. Koohbanani, B. Unnikrishnan, S. A. Khurram, P. Krishnaswamy, and N. Rajpoot, "Self-path: Self-supervision for classification of pathology images with limited annotations," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2845–2856, Oct. 2021.

[A26] F. Haghighi, M. R. H. Taher, Z. Zhou, M. B. Gotway, and J. Liang, "Transferable visual words: Exploiting the semantics of anatomical patterns for self-supervised learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2857–2868, Oct. 2021.

[A27] A. Mobiny *et al.*, "Memory-augmented capsule network for adaptable lung nodule classification," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2869–2879, Oct. 2021.

[A28] F. Xing, T. C. Cornish, T. D. Bennett, and D. Ghosh, "Bidirectional mapping-based domain adaptation for nucleus detection in cross-modality microscopy images," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2880–2896, Oct. 2021.

[A29] A. Tomczak *et al.*, "Multi-task multi-domain learning for digital staining and classification of leukocytes," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2897–2910, Oct. 2021.

[A30] L. Ju, X. Wang, X. Zhao, P. Bonnington, T. Drummond, and Z. Ge, "Leveraging regular fundus images for training UWF fundus diagnosis models via adversarial learning and pseudo-labeling," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2911–2925, Oct. 2021.

[A31] D. Tomar, M. Lortkipanidze, G. Vray, B. Bozorgtabar, and J.-P. Thiran, "Self-attentive spatial adaptive normalization for cross-modality domain adaptation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2926–2938, Oct. 2021.

[A32] S. Koehler *et al.*, "Unsupervised domain adaptation from axial to short-axis multi-slice cardiac MR images by incorporating pretrained task networks," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2939–2953, Oct. 2021.

References

[1] Z. Zhou, J. Y. Shin, S. R. Gurudu, M. B. Gotway, and J. Liang, "Active, continual fine tuning of convolutional neural networks for reducing annotation efforts," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 101997.

[2] Z. Zhou, J. Shin, R. Feng, R. T. Hurst, C. B. Kendall, and J. Liang, "Integrating active learning and transfer learning for carotid intima-media thickness video interpretation," *J. Digit. Imag.*, vol. 32, no. 2, pp. 290–299, Apr. 2019.

[3] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 17, 2021, doi: 10.1109/TPAMI.2021.3059968.

[4] M. M. R. Siddiquee *et al.*, "Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 191–200.

[5] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," 2018, *arXiv:1803.07728*. [Online]. Available: http://arxiv.org/abs/1803.07728

[6] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.

[7] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020, *arXiv:2002.05709*. [Online]. Available: http://arxiv.org/abs/2002.05709

[8] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Self-supervised learning for medical image analysis using image context restoration," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101539.

[9] Z. Zhou *et al.*, "Models genesis: Generic autodidactic models for 3D medical image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 384–393.

[10] Z. Zhou, V. Sodha, J. Pang, M. B. Gotway, and J. Liang, "Models genesis," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101840.

[11] S. Chakraborty, V. Balasubramanian, Q. Sun, S. Panchanathan, and J. Ye, "Active batch selection via convex relaxations with guaranteed solution bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 1945–1958, Oct. 2015.

[12] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7340–7351.

[13] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," 2021, *arXiv:2104.00298*. [Online]. Available: http://arxiv.org/abs/2104.00298

[14] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*. [Online]. Available: http://arxiv.org/abs/2010.11929

[15] M. Caron *et al.*, "Emerging properties in self-supervised vision transformers," 2021, *arXiv:2104.14294*. [Online]. Available: http://arxiv.org/abs/2104.14294

[16] Z. Zhou, "Towards annotation-efficient deep learning for computer-aided diagnosis," Ph.D. dissertation, Dept. Biomed. Inform. Program, Arizona State Univ., Tempe, AZ, USA, 2021.

[17] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[18] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[19] A.-A.-Z. Imran, "From fully-supervised, single-task to scarcely-supervised, multi-task deep learning for medical image analysis," Ph.D. dissertation, Dept. Comput. Sci., Univ. California, Los Angeles, CA, USA, 2021.

[20] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[21] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.

[22] N. Rieke *et al.*, "The future of digital health with federated learning," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–7, Dec. 2020.

[23] D. Yang *et al.*, "Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 101992.

[24] B. van Berlo, A. Saeed, and T. Ozcelebi, "Towards federated unsupervised representation learning," in *Proc. 3rd ACM Int. Workshop Edge Syst., Anal. Netw.*, Apr. 2020, pp. 31–36.

[25] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[26] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[27] Y. LeCun and I. Misra. *Self-Supervised Learning: The Dark Matter of Intelligence*. Accessed: Jun. 15, 2021. [Online]. Available: https://ai.facebook.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/

[28] F. Haghighi, M. R. H. Taher, Z. Zhou, M. B. Gotway, and J. Liang, "Learning semantics-enriched representation via self-discovery, self-classification, and self-restoration," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 137–147.

[29] R. Feng, Z. Zhou, M. B. Gotway, and J. Liang, "Parts2Whole: Self-supervised contrastive learning via reconstruction," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*. Cham, Switzerland: Springer, 2020, pp. 85–95.

[30] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*. [Online]. Available: http://arxiv.org/abs/1503.02531

[31] N. Tajbakhsh, M. B. Gotway, and J. Liang, "Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 62–69.

[32] N. Tajbakhsh, J. Y. Shin, M. B. Gotway, and J. Liang, "Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101541.

[33] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "A comprehensive computer-aided polyp detection system for colonoscopy videos," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, 2015, pp. 327–338.

[34] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 630–644, Feb. 2016.

[35] J. Y. Shin, N. Tajbakhsh, R. T. Hurst, C. B. Kendall, and J. Liang, "Automating carotid intima-media thickness video interpretation with convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2526–2535.

[36] N. Tajbakhsh, J. Y. Shin, R. T. Hurst, C. B. Kendall, and J. Liang, "Automatic interpretation of carotid intima–media thickness videos using convolutional neural networks," in *Deep Learning for Medical Image Analysis* Amsterdam, The Netherlands: Elsevier, 2017, pp. 105–131.