Addressing the Exorbitant Cost of Labeling Medical Images with Active Learning

Saba Rahimi, Ozan Oktay, Javier Alvarez-Valle, Sujeeth Bharadwaj

***Abstract*—**Successful application of deep learning in medical image analysis necessitates unprecedented amounts of labeled training data. Unlike conventional 2D applications, radiological images can be three dimensional (e.g. CT, MRI) consisting of many instances within each image. The problem is exacerbated when expert annotations are required for effective pixel-wise labeling, which incurs exorbitant labeling effort and cost. Active Learning is an established research domain that aims to reduce labeling workload by prioritizing a subset of informative unlabeled examples to annotate. Our contribution is a cost-effective approach for U-Net 3D models that uses Monte Carlo sampling to analyze pixel-wise uncertainty. Experiments on the AAPM 2017 lung CT segmentation challenge dataset show that our proposed framework can achieve promising segmentation results by using only 42% of the training data.

***Keywords*—**Image Segmentation, Active Learning, Convolutional Neural Network, 3D U-Net.

I. Introduction

With the widespread success of supervised deep learning on semantic image segmentation [1][2][3], the amount of labeled data required to train these networks has grown significantly. This has led to posing severe constraints on the applicability of deep neural networks in medical image analysis where only trained experts or qualified professionals can annotate data, and the costs associated with annotating sufficient examples is drastically high. Moreover, manual inspection of medical images can be very tedious and time-consuming [4], clinical experts have limited availability, and the imaging interpretation is subject to the experience of the specialist [5]. The problem is exacerbated in medical image semantic segmentation tasks, where effective pixel-level labeling is required. This challenge has resulted in the generation of significantly small public labeled datasets in the medical imaging field (30 for ISBI EM Challenge [6] and 85 for Gland Segmentation Challenge Contest in MICCAI 2015 [7]).

To reduce the costs associated with manual annotation and enlarge the training datasets, a number of techniques based on active learning have been proposed [8][9][10][11]. The core idea is that the framework, iteratively, selects samples to be labeled next, leading to a system that could potentially learn from only a fraction of the data. In each iteration, the framework selects a subset of samples from a large collection of unlabeled images according to a policy and queries their labels. Once labeled, the new candidates are added to the training set and the training model is fine-tuned using the augmented training set. This process is then repeated, with the annotated samples increasing in size over time, until the performance on a validation set plateaus. A thorough review of literature on active learning solutions for medical image segmentation can be found in Tajbakhsh et al. [12], with different active learning methods that select candidates for annotation based on informativeness and diversity of the data.

Existing uncertainty sampling techniques almost exclusively operate on 2D images and do not directly extend to many medical imaging modalities such as CT and MRI, which are inherently three-dimensional. To our knowledge, active learning applied to 3D U-Nets has not been investigated extensively. The work of [13] comes closest to ours in spirit, wherein the authors investigate active learning on whole 3D images; most other approaches to date have considered 2D images [14][15][16].

In this work, we evaluate the utility of active learning on 3D medical imaging data from the AAPM 2017 lung CT segmentation challenge [17]. We focus on 3D U-Net based segmentation models [18], which have demonstrated state-of-the-art results on various segmentation tasks in medical imaging and widely considered to be the de facto approach to semantic segmentation [19][20][21][22][23]. By introducing active learning with precise pixel-level uncertainty measurements on a publicly available dataset, our primary objective is to encourage increased adoption of active learning techniques within the medical imaging community.

II. Methods

Fig. 1 shows an overview of the proposed approach. Our proposed methodology consists of two major components including network training on the initial set of labeled data and image uncertainty scoring on the unlabeled data. The following section describes them in more detail:

Network Training: We choose 3D U-Nets as the architecture for segmentation, which remains one of the most popular convolutional neural networks (CNNs) in medical imaging

Saba Rahimi is with the Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada (e-mail: saba.rahimi@mail.utoronto.ca).

Ozan Oktay and Javier Alvarez-Valle are with Health Intelligence, Microsoft Research, Cambridge, CB1 2FB United Kingdom.

Sujeeth Bharadwaj is with AI & Advanced Architectures, Microsoft, Sunnyvale, CA 94085, United States.

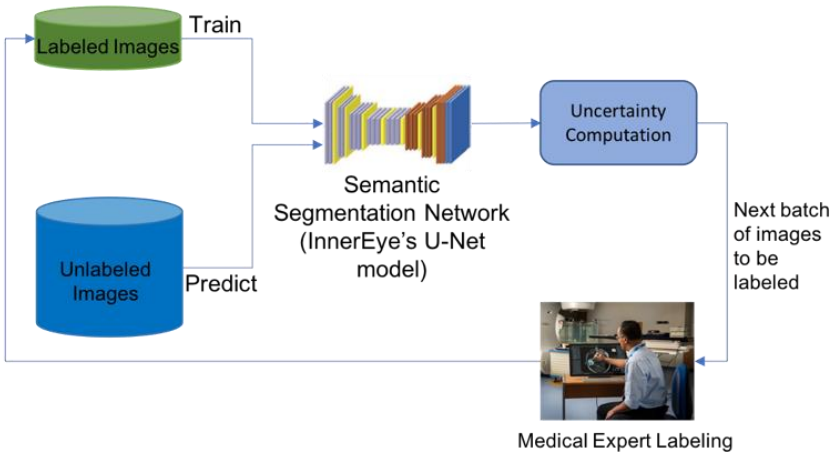

Fig. 1 Overview of the proposed method.

[18]. We first train the segmentation network using a small subset of training images that are randomly selected from the labeled dataset D^L, for which ground-truth annotation has been acquired.

Image Uncertainty Scoring: Next, our active learning method aims to query highly uncertain samples from the unlabeled dataset, D^U. We utilize a sample selection policy based on Monte Carlo dropout. Implementing dropout has been shown as an effective method to prevent overfitting during training [9], while at test time, it enables us to understand pixel-wise model uncertainty [14]. We then estimate the uncertainty of an individual image pixel by computing the variance of T different predictions generated by Monte Carlo sampling with dropout. The precision at pixel-wise uncertainty scores increases with increasing the number of dropout steps, T.

For a 2D image of [X×Y] pixels, we denote with $P_{(i,t)}^{(X,Y)}(c)$ our network's predicted probability for pixel *(i)* belonging to class *(c)* on run number *t*. Given this matrix of size [TxC] for each pixel, we use mean values of the variance of these values as an "uncertainty score":

$$V_i^{(X,Y)} = \frac{1}{C}\sum_{c=1}^{C} Var(P_{(i,t)}^{(X,Y)}(c)) \quad (1)$$

We then sum up all the pixel uncertainty values for the whole 3D image, obtaining a numerical score for each patient to estimate the prediction confidence, where a higher score is associated with the most uncertain segmentations. Assigning a numerical uncertainty score for each image has been previously integrated in the cost effective active learning (CEAL) sample selection method [24].

Next, the ground-truth annotations are collected for the selected unlabeled images, added to the labeled dataset and deleted from the unlabeled dataset. The segmentation network is then retrained on the updated training dataset. The active selection iteration is repeated until the annotation budget is exhausted or the entire dataset is labeled.

III. Results

All experiments were conducted using the publicly available AAPM 2017 lung CT segmentation challenge [17] which contains thoracic CT scans acquired from 60 patients. The dataset contains free form dense annotations for 5 organs: spinal cord, right lung, left lung, heart, and esophagus. The detailed descriptions of the dataset can be found in [25].

We use 50 random images from the challenge dataset, splitting it into 22 images for training, 22 for testing, and 6 for validation set. We start by training our network on 11 randomly selected images, then on each iteration we compute the uncertainty scores for each pixel, sum them up and select 5 images with the highest total uncertainties. We then estimate the uncertainty of an individual image pixel by running inference T = 5 times with dropout probability, p = 0.5. We construct a matrix of size [5 x 6] for each pixel, where C=6 is

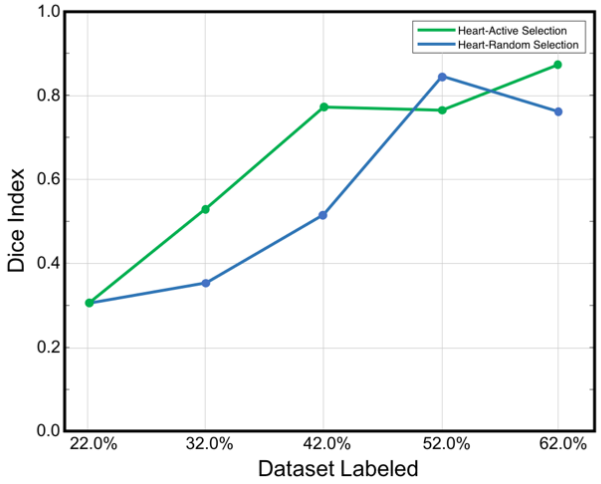

Fig. 2 Segmentation results using active learning compared to random selection.

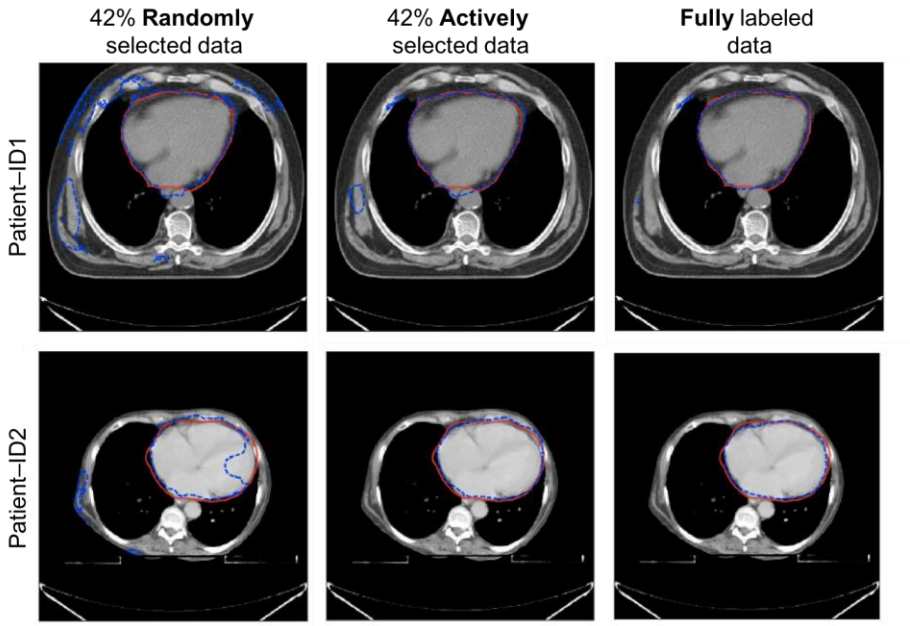

Fig. 3 Qualitative results of segmentation for two different patients. Red contour is the manual segmentation and blue contour is the UNet generated segmentation with random selection or active selection. The results of training with fully labeled dataset is shown for comparison.

the number of classes (5 organs + background) in the dataset. We assess the segmentation quality with Dice Coefficient [3] which is presented in:

$$Dice\ (A,B) = \frac{2|A\hat{B}|}{|A|+|B|} \tag{2}$$

where A represents the U-Net predicted masks for image and B its ground truth manual mask.

Fig. 2 shows the model's performance in segmenting the heart over multiple iterations of active learning. After running two active learning iterations with 24 training epochs per run, it already achieves a dice coefficient of 0.77 and it eventually outperforms random selection over 4 iterations by 0.11 points. We achieve validation dice coefficient of 0.77 on the second cycle when applying active learning sample selection, while training the same network on random sample selection results in dice coefficient of 0.52. This outperforms random by 0.25 points. We can see that active learning selection method outperforms random selection on the first cycle by 0.15 points. Active learning selection method outperforms random selection by 0.11 on the last cycle, reaching a mean dice coefficient of 0.87. This implies that annotating 42% of the objects in the images can already result in satisfactory segmentations, i.e., the annotation effort can be halved with no loss in performance. Fig. 3 shows our qualitative results where the result of random selection is compared to active selection for segmenting the heart of two different patients.

In this study, we proposed an active learning-based method for training 3D segmentation models using the U-Net architecture, which has been relatively underexplored in the medical imaging literature. Our results on the AAPM lung segmentation challenge dataset confirm that comparable accuracies can be achieved by actively labeling less than half of the entire training dataset. Since 3D medical images require slice-level annotations by trained experts, our results are a step towards reducing the exorbitant costs of labeling medical images. Future directions include active slice selection based on similar uncertainty measures as well as empirical validation of our approach on a number of other publicly available datasets.

ACKNOWLEDGMENT

This research was conducted as part of the first author's internship at Microsoft, Sunnyvale, CA. We would like to thank the Microsoft AI & Advanced Architectures group and Project InnerEye for computational resources, flexible codebase, and the many insightful discussions. The authors would like to thank M. Mesmakhosrowshahi and S. Bannur for their assistance.

REFERENCES

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

[2] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017.

[3] O. Ronneberger, P. Fischer, and T. Brox "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[4] W. P. Segars *et al.*, "Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization," *Med. Phys.*, vol. 40, no. 4, pp. 1–11, 2013.

[5] O. Oktay *et al.*, "Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers," *JAMA Netw. open*, vol. 3, no. 11, p. e2027426, 2020.

[6] I. A. Carreras *et al.*, “Crowdsourcing the creation of image segmentation algorithms for connectomics,” *Front. Neuroanat.*, vol. 9, no. November, pp. 1–13, 2015.
[7] K. Sirinukunwattana *et al.*, “Gland segmentation in colon histology images: The glas challenge contest,” *Med. Image Anal.*, vol. 35, pp. 489–502, 2017.
[8] Y. Fu, X. Zhu, and B. Li, “A survey on instance selection for active learning,” *Knowl. Inf. Syst.*, vol. 35, no. 2, pp. 249–283, 2013.
[9] Y. Gal, R. Islam, and Z. Ghahramani, “Deep Bayesian active learning with image data,” *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 3, pp. 1923–1932, 2017.
[10] W. H. Beluch, T. Genewein, A. Nürnberger, and J. M. Köhler, “The Power of Ensembles for Active Learning in Image Classification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 9368–9377, 2018.
[11] Y. Siddiqui, J. Valentin, and M. Nießner, “Viewal: Active learning with viewpoint entropy for semantic segmentation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 9430–9440, 2020.
[12] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, “Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation,” *Med. Image Anal.*, vol. 63, p. 101693, 2020.
[13] W. Kuo, C. Häne, E. Yuh, P. Mukherjee, and J. Malik, “Cost-sensitive active learning for intracranial hemorrhage detection,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11072 LNCS, pp. 715–723, 2018.
[14] M. Gorriz, X. Giro-I-Nieto, A. Carlier, and E. Faure, “Cost-effective active learning for melanoma segmentation,” *arXiv*, no. Nips, 2017.
[15] L. Yang, Y. Zhang, J. Chen, S. Zhang, and D. Z. Chen, “Suggestive annotation: A deep active learning framework for biomedical image segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10435 LNCS, no. 1, pp. 399–407, 2017.
[16] D. Mahapatra, B. Bozorgtabar, J. P. Thiran, and M. Reyes, “Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11071 LNCS, pp. 580–588, 2018.
[17] L. Vinet and A. Zhedanov, “A ‘missing’ family of classical orthogonal polynomials,” *Journal of Physics A: Mathematical and Theoretical*, 2011. .
[18] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-net: Learning dense volumetric segmentation from sparse annotation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 424–432, 2016.
[19] S. Chen and M. De Bruijne, “An End-to-end Approach to Semantic Segmentation with 3D CNN and Posterior-CRF in Medical Images,” *arXiv*, no. Nips, pp. 3–6, 2018.
[20] J. De Fauw *et al.*, “Clinically applicable deep learning for diagnosis and referral in retinal disease,” *Nat. Med.*, vol. 24, no. 9, pp. 1342–1350, 2018.
[21] T. Falk *et al.*, “U-Net: deep learning for cell counting, detection, and morphometry,” *Nat. Methods*, vol. 16, no. 1, pp. 67–70, 2019.
[22] Y. Zhou, S. Bai, C. Wang, X. Chen, E. Fishman, and A. L. Yuille, “Prior-aware Neural Network for Partially-Supervised Multi-Organ Segmentation,” pp. 10672–10681.
[23] H. Chen, Q. Dou, L. Yu, J. Qin, and P. A. Heng, “VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images,” *Neuroimage*, vol. 170, no. April 2017, pp. 446–455, 2018.
[24] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, “Cost-Effective Active Learning for Deep Image Classification,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 12, pp. 2591–2600, 2017.
[25] J. Yang *et al.*, “Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017,” *Med. Phys.*, vol. 45, no. 10, pp. 4568–4581, 2018.