Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Doctoral Consortium Track

Hybrid Learning System for Large-scale Medical Image Analysis

Zehua Cheng and **Lianlong Wu**
University of Oxford
zehua.cheng@cs.ox.ac.uk, lianlong.wu@cs.ox.ac.uk

Abstract

Adequate annotated data cannot always be satisfied in medical imaging applications. To address such a challenge, we would explore ways to reduce the quality and quantity of annotations requirements of the deep learning model by developing a hybrid learning system. We combined self-supervised, semi-supervised, and weak-supervised learning to improve annotation utilization. Our primary research work on 2D medical image detection under poor annotation conditions has found that better regularization and adversarial loss can improve the robustness and performance with poor annotation conditions.

1 Introduction

Reliable annotation is a critical necessity for building a deep learning pipeline. However, due to human resources constraints, high-quality annotation cannot always be accommodated. Such constraints have become obstacles to applying deep learning for large-scale medical image analysis. To solve such a challenge, we need to lower the annotation quality and quantity requirements for the existing deep learning models in medical image analysis.

We will progressively reduce annotation requirements to achieve large-scale medical imaging analysis under clinical conditions. In this case, we would develop a hybrid learning system to utilize the large-scale unlabelled medical images and labelled medical images with unknown annotation quality. The annotation could be fine-grained annotation like key point and segmentation mask or coarse annotation like classification label or inexact bounding box or segmentation mask.

2 Current Works

2.1 Robust Medical Detection with Insufficient Annotations

Manual annotations are not always exact due to time, cost, and experience constraints in clinical practice. Such cases would produce inexact bounding box annotations where the bounding box is either too big for the target or partially covered. We evaluated different detection models over the different medical detection benchmarks with different levels of detection annotation quality. We present part of the experimental results in Table 1. We found performance and robustness degradation in all benchmarks when the bounding box annotation is not strictly covered the target. We developed a regularization policy to filter the positive samples using the centroid feature maps of the different levels of Feature pyramid networks (FPNs) [Lin *et al.*, 2017]. We further constructed an adversarial loss to utilize the side-product of the regularization to improve the model robustness.

Method	Baseline	Inaccurate Annotation
Faster R-CNN	67.3	54.3
Cascade R-CNN	66.4	55.4
RetinaNet	61.3	53.2

Table 1: Average precision (%) of different detection model with ResNet50 [He *et al.*, 2016] in SIIM-FISABIO-RSNA COVID-19 Detection benchmark [Lakhani and others, 2021] with baseline (normal annotation) and inaccurate annotation.

We believe this is a good start to lower the annotation requirements in medical image analysis. Since FPNs are also applied in segmentation tasks, we can also apply such a method over the segmentation tasks and integrate to a wide range of applications.

2.2 Hybrid Learning System for Large-scale Medical Image Analysis

Currently, there are several new training paradigms for deep learning and unsupervised and supervised learning. *Self-supervised learning* aim to obtain a powerful pre-trained model by training pretext tasks and applying transfer learning to downstream tasks. Typically, pretext tasks do not require extensive human efforts, and the majority of tasks aim to design a reconstruction pipeline to extract the associated features. [Jing and Tian, 2020]. *Semi-supervised learning* refers to learning with limited annotated in conjunction with many unlabelled data [Van Engelen and Hoos, 2020]. *Weak-supervised learning* tried to learn a model with coarse-grained annotations [Zhou, 2018]. Such learning paradigms reduce the annotation requirement from a model training perspective. Each learning paradigm, however, has its own limitations. Self-supervised learning may require additional time to train a pre-trained model, making it inefficient for large-scale applications. A detailed guidelines [Oliver *et al.*, 2018] has presented theoretical and empirical proof to support the claim that unlabelled data can **always** improve the performance of the deep networks in image classification when the annotation is correctly assigned. Large-scale weak-supervised learning may harmed localization performance when transferring to downstream tasks [Mahajan, 2018].

In this case, we would like to develop a system that merges the existing training pipeline to solve the large-scale medical image analysis with different types of annotations under unknown annotation quality.

Primary attempt apply self-supervised learning on large-scale medical images shows large-scale medical images can improve the model performance and robustness [Ghesu, 2022]. We would extend such framework by merging different types of learning system. During our primary exploration on the medical images, we have found semi-supervised learning could serve as a bridge to self-supervised learning and weak-supervised learning. Our preliminary experiments found that applying semi-supervised learning to the downstream task of self-supervised learning improves overall performance and mitigates the sacrifice of localization performance in large-scale weakly supervised learning.

2.3 Improved Hybrid Learning System with Graph Constraints

We would introduce graph constraints to learning pipeline to improve the hybrid learning system that discussed in the Section 2.2. Graph constrain has widely applied to learn the imbalanced data [Du *et al.*, 2021]. A self-supervised learning pipeline could also benefit from the graph constraints [Li *et al.*, 2021] to improve the robustness in learning fine-grained visual features. Thus, by introducing graphical constraints, information can be utilized in a more organized way, rather than being driven entirely by gradients.

Meanwhile, we found that knowledge graphs are heavily used in the field of medical diagnosis [Li *et al.*, 2019]. Introducing knowledge graph could also improve the annotation utilization and the model performance. [Xie *et al.*, 2022] introduced the graph neural networks to process the relationship of the predicted mask for instance segmentation task. Therefore, applying knowledge graph to the hybrid learning system is a prominent option. Knowledge graphs not only model the domain knowledge explicitly, but also incorporate prior exact expert rules or constraints. In particular, we would investigate using symbolic rule learner [Wu *et al.*, 2022] to conduct inductive relational learning, in order to enhance the precision effectively and efficiently.

3 Future Works

Based on our primary exploration, while solutions in healthcare have some domain-specific requirements, the solution for challenges mentioned in the above section are broad-spectrum and not limited to the medical domain. We would further explore the potential of our hybrid learning system over the large-scale data and cooperate with medical institutions to further verify our design.

Acknowledgements

Zehua Cheng is under supervision of Prof. Thomas Lukasiewicz. Lianlong Wu is under supervision of Prof. Emanuel Sallinger and Prof. Georg Gottlob.

References

[Du *et al.*, 2021] Guodong Du, Jia Zhang, et al. Graph-based class-imbalance learning with label enhancement. *IEEE TNNLS*, 2021.

[Ghesu, 2022] Florin et al. Ghesu. Self-supervised learning from 100 million medical images. *arXiv preprint arXiv:2201.01283*, 2022.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[Jing and Tian, 2020] Longlong Jing and Yingli Tian. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE TPAMI*, 43(11):4037–4058, 2020.

[Lakhani and others, 2021] Paras Lakhani et al. The 2021 SIIM-FISABIO-RSNA Machine Learning COVID-19 Challenge: Annotation and standard exam classification of COVID-19 chest radiographs. pages 1–25, 2021.

[Li *et al.*, 2019] Christy Y Li, Xiaodan Liang, Zhiting Hu, and Eric P Xing. Knowledge-driven encode, retrieve, paraphrase for medical image report generation. In *AAAI*, volume 33, pages 6666–6673, 2019.

[Li *et al.*, 2021] Junnan Li, Caiming Xiong, and Steven CH Hoi. Comatch: Semi-supervised learning with contrastive graph regularization. In *ICCV*, pages 9475–9484, 2021.

[Lin *et al.*, 2017] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of CVPR*, pages 2117–2125, 2017.

[Mahajan, 2018] Dhruv et al Mahajan. Exploring the limits of weakly supervised pretraining. In *ECCV*, pages 181–196, 2018.

[Oliver *et al.*, 2018] Avital Oliver, Augustus Odena, et al. Realistic evaluation of deep semi-supervised learning algorithms. *NeurIPS*, 31, 2018.

[Van Engelen and Hoos, 2020] Jesper E Van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine Learning*, 109(2):373–440, 2020.

[Wu *et al.*, 2022] Lianlong Wu, Emanuel Sallinger, Evgeny Sherkhonov, Sahar Vahdati, and Georg Gottlob. Rule learning over knowledge graphs with genetic logic programming. In *ICDE*, pages 1–13, 2022.

[Xie *et al.*, 2022] Chris Xie, Arsalan Mousavian, Yu Xiang, and Dieter Fox. Rice: Refining instance masks in cluttered environments with graph neural networks. In *Conference on Robot Learning*, pages 1655–1665. PMLR, 2022.

[Zhou, 2018] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National science review*, 5(1):44–53, 2018.