A Comparative Evaluation of Deep Learning Based Region-of Interest Detection of Images in Biomedical Literature *

Md Mahmudur Rahman, *Member, IEEE* and Bikesh Regmi

Abstract— **Biomedical images are frequently used in articles to illustrate medical concepts and highlight regions-of-interests (ROIs). In many cases multiple annotation markers such as different arrows, letters or symbols overlaid on figures are often pointing to different Region-of-Interests (ROIs) relevant to the article. This work presents a proof-of-concept based comparative evaluation of three popular Deep Learning based object detection techniques (Mask R-CNN, YOLOv5, and YOLOv7) to identify such ROIs in a dataset of 450 Chest CT images which are appeared in biomedical articles. The results demonstrate that all three DL-based object detection techniques achieved high accuracy in recognizing and localizing the ROIs in biomedical images, whereas YOLOv7 exhibited the highest precision of 92.5%, indicating its ability to accurately identify ROIs.**

Keywords—biomedical images, region-of-interest, object detection, classification, evaluation

I. Introduction

Due to the ongoing progress in biomedical domain, a wide variety of users, such as patients, researchers, general practitioners, and clinicians often use tools to search for relevant and actionable information from biomedical literature to their individual needs. It creates the need of literature-based informatics for managing the rapidly increasing volume of information in the biomedical domain [1]. For example, according to PubMed Central® (PMC), a free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM) alone contains more than 8 million articles with an average 3-5 figures as of 2022 [2]. Authors of biomedical articles often use arrows, pointers, and other annotations such as text labels overlaid on figures and illustrations in the biomedical articles to highlight significant areas as Region-of-Interests (ROIs). These annotations are then referenced and correlated with concepts in the caption text or figure citations in the article text. The pointers/markers are good source of information to locate specific visual patterns as ROIs in the image [9]. For example, Fig. 1 shows different arrowheads pointing to "***honeycomb'*** (hexagonal wax cells like structure), a black arrow pointing to ***"cyst"*** and a white arrow pointing to **"bronchitis"** (diffuse thickening of the airway walls giving the appearance of thick lines and rings throughout the lungs) patterns in a chest CT image (along with caption at the top) as depicted in a biomedical article.

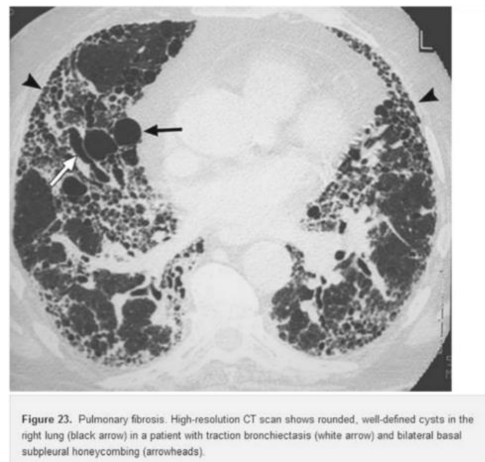

Fig. 1 Example chest CT image with associated caption and different arrows/pointers

These annotation markers can assist in extracting relevant ROIs (Fig. 2) within the image that are likely to be highly relevant to the discussion in the article text.

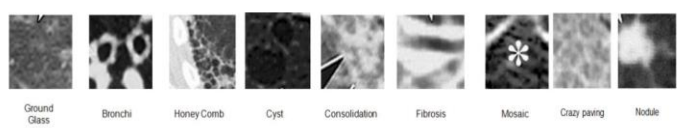

Fig. 2 Example ROIs in the thoracic imaging glossary

Image regions can then be annotated using biomedical concepts from extracted snippets of text in captions that might be further identified using existing textual ontologies, such as the Unified Medical Language System (UMLS). Such a resource could assist in reducing the semantic gap problem in image classification and retrieval. However, these images are always in low-resolution compared to their clinical counterparts, in varying sizes and lightning conditions and moreover the dataset is highly imbalanced where a few concept categories (patterns) occur more frequently compared to other less frequent ones.

This work presents a proof-of-concept experiment by annotating 450 thoracic images of chest images for object detection using LabelImg [3] as graphical image annotation tool. The images are appeared in biomedical articles, which are obtained through a large benchmark collection under

*Resrach supported by NSF.

Md Mahmudur Rahman is with the Computer Science Department, Morgan State University, Baltimore, MD, 21251 USA (phone: 443-885-1056; e-mail: md.rahman@ morgan.edu).

Bikesh Regmi is with the Computer Science Department, Morgan State University, Baltimore, MD, 21251 USA (e-mail: bireg1@ morgan.edu).

ImageCLEFmed [4]. The main objective of this work is to evaluate the effectiveness of DL-based object detection models in localizing ROIs in such biomedical images.

II. DL-BASED OBJECT DETECTION

For the past several years, DL based algorithms have shown great promise in improving the accuracy of object detection in recent years [5]. One of the pioneer methods is the Region-Based Convolutional Neural Networks (R-CNNs) first introduced around 2012, which are a family of techniques (e.g., Fast R-CNN, Faster R-CNN, Mask R-CNN) for addressing object localization and recognition tasks, which operate in two stages [6]. Mask R-CNN extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition [7]. On the other hand, YOLO (You Only Look Once), is a second family of techniques designed for speed and real-time use with state-of-the-art accuracy in recent years [8,9]. In these techniques, there is no region creation and then again processing on top of that, rather there is one convolution network that creates boxes and class predictions for each box as one stage detector.

III. EXPERIMENTS AND RESULTS

We evaluated the performance of Mask R-CNN, YOLOv5, and YOLOv7 by measuring their accuracy in detecting and localizing the annotated markers within the Chest CT images. The evaluation metrics used were ***precision, recall***, and ***mean average precision (mAP)***. Additionally, we compared the computational efficiency of the techniques to assess their practical variability.

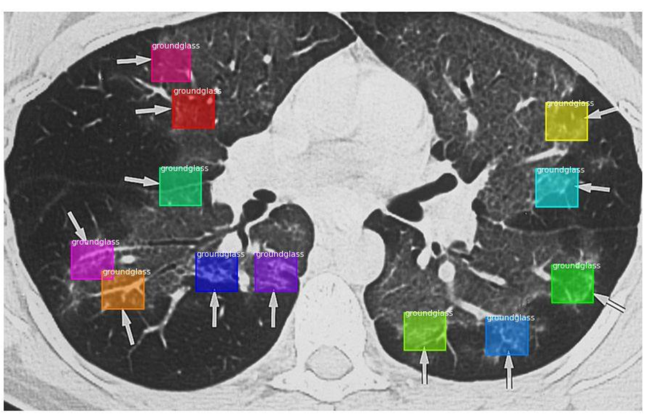

Fig. 3 Sample result of ROI detection as 'ground glass' pattern

The results in Table I demonstrate that all three DL-based object detection techniques achieved high accuracies in recognizing and localizing (annotation markers) the ROIs in biomedical images. YOLOv7 exhibited the highest precision of 92.5%, indicating its ability to accurately identify ROIs. YOLOv5 and Mask R-CNN also demonstrated respectable precision scores of 87.2% and 85.6%, respectively. In terms of recall, Mask R-CNN achieved 87.3%, indicating its effectiveness in capturing a high percentage of the annotated markers. YOLO v5 and YOLOv7 showed slightly lower recall rates of 84.6% and 81.8%, respectively, but still performed well in detecting the markers. Fig. 3 shows an example image where the ROIs are detected correctly as "ground glass" patterns which are pointed by white arrows.

TABLE I. ACCURACY COMPARISON OF DIFFERENT DL-BASED OBJECT DETECTION METHODS

Method	Precision (%)	Recall (%)	mAP (%)
Mask R-CNN	85.6	81.8	0.83
YOLOv5	87.2	84.6	0.85
YOLOv7	**92.5**	**87.3**	**0.89**

Regarding computational efficiency, YOLOv5 and YOLOv7 exhibited faster inference times compared to Mask R-CNN. This makes them more suitable for real-time applications where speed is crucial, although with a slight compromise on precision and recall rates. In conclusion, the results of this study demonstrate that DL-based object detection techniques, specifically Mask R-CNN, YOLOv5, and YOLOv7, can achieve high accuracy in detecting and classifying different types of ROIs in low-resolution chest CT images as often appeared in biomedical journal articles.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation (NSF) grant (ID. 2131307) under CISE-MSI program.

REFERENCES

[1] M. S. Simpson, D. You, M. M. Rahman, Z. Xue, D. Demner-Fushman, S. K. Antani and G. R. Thoma G., "Literature-based biomedical image classification and retrieval, " Comput Med Imaging Graph. Vol. 39, pp. 3-13, 2015

[2] Z. Lu. "Pubmed and beyond: a survey of web tools for searching biomedical literature", Vol. baq036, 2011

[3] LabelImg for Labeling Object Detection Data: https://blog.roboflow.com/labelimg/

[4] H. Müller, A. Herrera, J. Kalpathy-Cramer, D. Demner-Fushman, S. K. Antani and E. Ivan, "Overview of the ImageCLEF2012 Medical Image Retrieval and Classification Tasks," In: The Working Notes for the CLEF 2012 Labs and Workshop, Rome, Italy, 17–20 Sept 2012

[5] Z. Q. Zhao, P. Zheng, S. -T. Xu and X. Wu, "Object Detection With Deep Learning: A Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[6] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.

[7] He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. IEEE Trans Pattern Anal Mach Intell. 2020 Feb;42(2):386-397. doi: 10.1109/TPAMI.2018.2844175. Epub 2018 Jun 5. PMID: 29994331.

[8] Object Detection using YOLOv5 OpenCV DNN in C++ and Python: https://learnopencv.com/object-detection-using-yolov5-and-opencv-dnn-in-c-and-python/

[9] C. Wang, A. Bochkovskiy, and H. M. Liao, (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. 10.48550/arXiv.2207.026