DeepLesion: a Diverse and Large-scale Database of Significant Radiology Image Findings

Ke Yan, Xiaosong Wang, Le Lu, Ling Zhang,
Mohammadhadi Bagheri, Ronald M. Summers

Imaging Biomarkers and Computer-Aided Diagnosis Laboratory
National Institutes of Health Clinical Center, Bethesda, MD 20892
{ke.yan, xiaosong.wang, ling.zhang3, mohammad.bagheri, rms}@nih.gov, lelu@cs.jhu.edu

1 Purpose

Large-scale datasets with diverse images and dense annotations are important in training effective computer-aided detection/diagnosis (CADe/CADx) algorithms, especially in the era of deep learning. However, most publicly available medical image datasets are either small (with hundreds of cases) [3] or contain only image-level labels. In addition, most datasets focus on only certain body parts, such as lung and liver [1]. Annotating medical image datasets is time-consuming and requires extensive clinical training. To solve these problems, we introduce our work in [5, 6]. A data mining strategy is proposed to harvest large-scale lesion-level annotations with minimum manual effort. DeepLesion, a lesion database with over 32K lesions of various types in CT images, was collected and used to train algorithms for multi-class lesion retrieval, matching, and detection. The dataset will be publicly released soon.

2 Method

Radiologists routinely mark and measure significant image findings according to the RECIST guidelines [2]. These annotations, sometimes called “bookmarks”, have been collected over years in hospitals’ picture archiving and communication systems (PACS). Without loss of generality, we study one type of bookmark in CT images: RECIST diameters, which are measurements of the lesions containing one long axis and one short axis. We collected these coordinates from the PACS server of NIH and obtained DeepLesion [6], which includes 32,120 axial CT slices from 10,594 studies of 4,427 unique patients. There are 1–3 lesions in each image, adding up to 32,735 lesions altogether. A bounding box was generated tightly around the two diameters of each lesion for further analysis. We also defined and partially labeled 8 lesion types: lung, mediastinum, liver, kidney, abdomen, pelvis, soft tissue, and bone.

DeepLesion can be utilized in various applications. First, we use it for lesion retrieval [6], which aims at finding similar lesions in the database for the query one to help understand it. To model the similarity relationship of lesions, we extract multiple supervision information including lesion types, normalized location coordinates, and sizes. Then, a triplet network with a sequential sampling strategy is proposed to learn an embedding vector for each lesion. Finally, similar lesions can be retrieved by searching the nearest neighbors in the embedding space. The second application is to match and track the same lesion instances of one patient across multiple longitudinal studies, which can be achieved by running a graph-based edge pruning algorithm on the learned lesion embeddings [6]. We also proposed 3D context enhanced region-based convolutional neural network (3DCE) for lesion detection [5]. 3DCE incorporates 3D context information efficiently by aggregating feature maps of multiple 2D image slices. It is easy to train and end-to-end in both training and inference.

3 Results

For lesion retrieval, the average retrieval error of the top-5 retrieved lesions on the test set of DeepLesion (4927 samples) were computed. We achieved 8.5%, 7.2%, and 5.1% error rates for lesion type, location, and size. For lesion matching, we manually grouped 1313 lesions from 103 patients in DeepLesion to 593 groups for evaluation. The accuracy is promising with an area under the curve (AUC) of 95.9%. DeepLesion is a challenging dataset for lesion detection since it contains all kinds of lesions and many lesions and nonlesions look similar. The sensitivity at 4 false positives per image of the popular faster RCNN [4] is 80.32%, while our 3DCE achieved 84.37% by leveraging the 3D context. Lung, mediastinum, and liver lesions have the highest detection accuracy, whereas bone and kidney lesions are the hardest to detect.

4 Conclusion

We propose to mine the bookmarks from PACS to acquire lesion-level annotations for medical image datasets. A large-scale and diverse dataset, DeepLesion, was collected and will be released. We expect it to be useful in applications such as lesion detection, retrieval, matching, segmentation, and classification. One limitation is that the dataset does not contain segmentation masks.

Acknowledgments: This research was supported by the Intramural Research Program of the NIH Clinical Center. We thank NVIDIA for the GPU card donation.

References

1. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The cancer imaging archive (TCIA): Maintaining and operating a public information repository. J. Digit. Imaging 26(6), 1045–1057 (2013)
2. Eisenhauer, E., Therasse, P., et al.: New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European journal of cancer 45(2), 228–247 (2009)
3. Kohli, M.D., Summers, R.M., Geis, J.R.: Medical Image Data and Datasets in the Era of Machine Learning - Whitepaper from the 2016 C-MIMI Meeting Dataset Session. J. Digit. Imaging 30(4), 392–399 (2017)
4. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: NIPS. pp. 91–99 (2015)
5. Yan, K., Bagheri, M., Summers, R.M.: 3D Context Enhanced Region-based Convolutional Neural Network for End-to-End Lesion Detection. In: MICCAI (2018)
6. Yan, K., Wang, X., Lu, L., Zhang, L., Harrison, A., Bagheri, M., Summers, R.: Deep Lesion Graphs in the Wild: Relationship Learning and Organization of Significant Radiology Image Findings in a Diverse Large-scale Lesion Database. In: CVPR (2018)