LESION SEARCH WITH SELF-SUPERVISED LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Content-based Image Retrieval (CBIR) with deep learning accelerates clinicians' interpretation of similar cases to improve diagnostics. Lacking diverse and expert-verified annotations, self-supervised learning (SSL) spurs its advantages in medical imaging. We build off a CBIR starting from a contrastive learning backbone (SimCLR) and add a Generalized-Mean (GeM) pooling layer followed by L2 normalization to classify lesion types and retrieve similar lesions before clinicians' analysis. Results have shown improved retrieval performances. We additionally develop an open-source application for image analysis and lesion retrieval. The application is easy to integrate, relieving manual retrieval efforts and suggesting the potential to support clinicians' everyday activities.

1 INTRODUCTION

Radiologists' interpretation of lesions (nodules, tumor-like, and surface content) requires careful analysis. However, lesion images employ a wide variety of sizes, shapes (e.g. elliptical, irregular), color (e.g. grayscale), and textures (e.g. convex, nodule, distortion) (Hofman & Hicks (2016)), which makes the manual work time-consuming and at risk of errors due to expensive annotations and visual fatigue. Additionally, through automatic retrieval such as content-based image retrieval (CBIR) is challenging because the same lesion type may not be visually similar. This can be due to intra-class variations, while different lesion types may appear similarly once images were taken at similar stages of disease progression.

In this work, we develop a self-supervised lesion feature extractor for clinical tasks, including image retrieval and type classification. The method is shown in Figure 1. We also built a web-based open-source DICOM (Digital Imaging and Communications in Medicine) viewer for remotely analyzing images, interactively annotating ROI queries, and retrieving highly similar lesions.

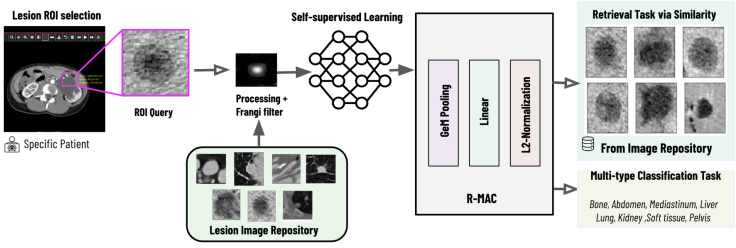

Figure 1: Pipeline for analyzing DICOM (Digital Imaging and Communications in Medicine) files, region selection, and lesion retrieval.

2 APPLICATION: DICOM VIEWER FOR USER INTERFACE

The web-accessible application consists of two components: UI front-end and CBIR back-end. AngularJS-based UI viewer presents toolbars for loading batch images, dragging, zooming, color shifting, annotating, and analyzing DICOM metadata for versatile functionalities, shown in Figure 2. The design supports a simple drag-and-drop of several DICOM files for either analyzing images in a batch or analyzing disease progression in a time frame. Annotations are stored in the JSON format and can either be saved locally or sent to CBIR back-end involving SSL models.

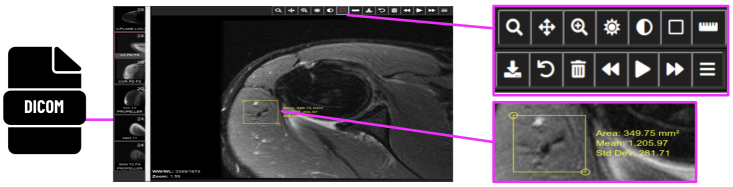

Figure 2: The front-end UI can load a multiple series of DICOM files. It offers a variety of functionalities such as annotations, size measures, contrasts, navigating image stacks, etc.

3 METHODS AND RESULTS

We used the DeepLesion dataset (Yan et al. (2018)) for analysis, details are in Appendix A. First, we applied the Hessian-based Frangi filter (Frangi et al. (1998); Kroon & Schrijver (2009); Longo et al. (2020)) to maximize lesion detection while minimizing noises. Details are in Appendix B. Second, CBIR ranks candidate descriptors C in ascending order according to their closer cosine distances to a query $\mathbf{Q}$. We choose top-9 candidates to cover enough similarity. Loss function and distance details are in Appendix C. Finally, SimCLR backbone adds a Generalized-Mean (GeM) pooling layer followed by L2 normalization to improve feature extraction. Experimental details are in Appendix D.

Performance of classification. All experiments start from two baselines: logistic regression and random forest. We compare our method, SimCLR-based, with same procedures on Variational Auto-encoder (VAE) backbone (Kingma & Welling (2013)). Table 1 summarizes results, showing that ours outperforms other methods, particularly the baselines by a large margin.

Table 1: Averaged classification performances measured within 10 runs

Model	Filter	Accuracy (%) (std.)	F1-score (%) (std.)
Logistic Regression	Frangi	30.02 ($\pm$0.47)	29.88 ($\pm$0.46)
Random Forest	Frangi	31.44 ($\pm$0.53)	30.80 ($\pm$0.35)
VAE-based	None	64.52 ($\pm$0.34)	63.40 ($\pm$0.29)
VAE-based	Frangi	**74.39** ($\pm$0.26)	**75.71** ($\pm$0.18)
Ours	None	69.24 ($\pm$0.62)	65.27 ($\pm$0.43)
Ours	Frangi	**80.39** ($\pm$0.34)	**80.93** ($\pm$0.21)
Azizi et al. (2021)	None	63.44 ($\pm$0.13)	-
Syrén Grönfelt (2021)	None	67.5	-
Yan et al. (2020)	None	**80.7**	-

Performance of CBIR. Table 2 presents the CBIR performances under different settings through the patient index-wise approach: 1) all-patient, 2) intra-patient, and 3) inter-patient.

A few visual samples of retrieved top-9 candidates are shown in Figure 3. Random queries are highlighted within pink boxes, and each row is a new search.

Table 2: CBIR performance under three settings: a) All-patient (all lesions can be candidates); b) Intra-patient (lesions from the same patient are candidates); c) Inter-patient (lesions from different patients are candidates).

Evaluation setting	mAP@10	Precision@1	Precision@3	Precision@5	Precision@10
All-patients	0.6759	0.9982	0.7336	0.6433	0.5416
Same-patient	0.6526	0.9992	0.7574	0.6330	0.4584
Cross-patient	0.4580	0.5027	0.4703	0.4578	0.4306

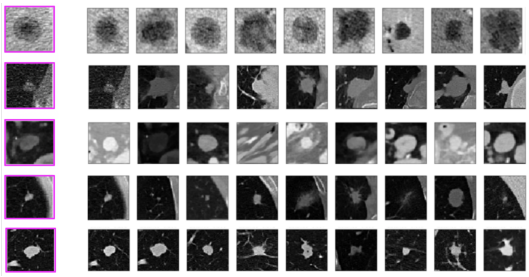

Figure 3: Each row is a separate search for a random query, followed by top-9 candidates on the right, ranked by cosine distances metric.

4 CONCLUSIONS

We present an interactive and open-source application that facilitates lesions analysis and retrievals based on self-supervised learning, It simplifies and improves clinicians' experiences in analyzing lesion variations. Experiments show that Frangi filter, GeM pooling and L2-normalization together can improve both intra- and inter-patient lesion retrieval.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Shekoofeh Azizi, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh, Alan Karthikesalingam, Simon Kornblith, Ting Chen, et al. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3478–3488, 2021.

Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International conference on medical image computing and computer-assisted intervention*, pp. 130–137. Springer, 1998.

R Gómez. Understanding ranking loss, contrastive loss, margin loss, triplet loss, hinge loss and all those confusing names. *Raúl Gómez blog*, pp. 12, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Michael S Hofman and Rodney J Hicks. How we read oncologic fdg pet/ct. *Cancer Imaging*, 16(1): 1–14, 2016.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013. URL `https://arxiv.org/abs/1312.6114`.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

Dirk-Jan Kroon and M Schrijver. Hessian based frangi vesselness filter. *MATLAB File Exchange*, 24409, 2009.

Antonia Longo, Stefan Morscher, Jaber Malekzadeh Najafababdi, Dominik Jüstel, Christian Zakian, and Vasilis Ntziachristos. Assessment of hessian-based frangi vesselness filter in optoacoustic imaging. *Photoacoustics*, 20:100200, 2020.

Tri-Cong Pham, Chi-Mai Luong, Muriel Visani, and Van-Dung Hoang. Deep cnn and data augmentation for skin lesion classification. In *Asian Conference on Intelligent Information and Database Systems*, pp. 573–582. Springer, 2018.

Natalie Syrén Grönfelt. Pretraining a neural network for hyperspectral images using self-supervised contrastive learning, 2021.

Mehreen Tariq, Sajid Iqbal, Hareem Ayesha, Ishaq Abbas, Khawaja Tehseen Ahmad, and Muhammad Farooq Khan Niazi. Medical image based breast cancer diagnosis: State of the art and future directions. *Expert Systems with Applications*, 167:114095, 2021.

Gijs van Tulder and Marleen de Bruijne. Combining generative and discriminative representation learning for lung ct analysis with convolutional restricted boltzmann machines. *IEEE transactions on medical imaging*, 35(5):1262–1272, 2016.

Ke Yan, Xiaosong Wang, Le Lu, and Ronald M Summers. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of medical imaging*, 5(3):036501, 2018.

Ke Yan, Jinzheng Cai, Dakai Jin, Shun Miao, Adam P Harrison, Dazhou Guo, Youbao Tang, Jing Xiao, Jingjing Lu, and Le Lu. Self-supervised learning of pixel-wise anatomical embeddings in radiological images. *arXiv preprint arXiv:2012.02383*, 2020.

Yang You, Igor Gitman, and Boris Ginsburg. Large batch training of convolutional networks. *arXiv preprint arXiv:1708.03888*, 2017.

A DATASET

DeepLesion Dataset. Previously, most algorithms considered one or a few types of lesions (Pham et al. (2018); Tariq et al. (2021); van Tulder & de Bruijne (2016)). Yan et al. developed the large and publicly available DeepLesion dataset from the NIH Clinical Center (Yan et al. (2018)) towards developing universal lesion detections. In clinical practice, DeepLesion dataset is widely accepted for monitoring cancer patients due to its recorded diameters in accordance with the Solid Tumor Response Evaluation Criteria (RECIST), which is one of the commonly used means of criteria. DeepLesion dataset consists of 33,688 PACS-bookmarked CT images from 10,825 studies of 4,477 unique patients (Yan et al. (2018)). It includes various lesions across the human body, such as lung, lymph, liver, etc. Four cardinal directions (left, top, right, bottom) enclose each lesion in every CT slice as a bounding box to mark coarse annotations with labels. We crop regions of interest (ROIs) based on these bounding boxes instructed in dataset to comprise excessive instances for model training.

B FRANGI FILTER FOR LESION DETECTION

Frangi filter function has two components and λ represents eigen values (Frangi et al. (1998)): if either $\lambda_2 > 0$ or $\lambda_3 > 0$, $V_f(\lambda) = 0$. Otherwise:

$$V_f(\lambda) = (1 - e^{-\frac{R_A^2}{2\alpha^2}}) \cdot e^{-\frac{R_B^2}{2\beta^2}} \cdot (1 - e^{-\frac{s^2}{2\gamma^2}}), \tag{1}$$

where $R_A = \frac{|\lambda_2|}{|\lambda_3|}$, $R_B = \frac{|\lambda_1|}{\sqrt{|\lambda_2 \cdot \lambda_3|}}$, and $s = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$. The modification enhances contours by restricting λ such that $|\lambda 1| \leq |\lambda_2| \leq |\lambda_3|$ at the scale "s". To suppress background noises that are not contours, we set λ_3 = 0 if $\lambda_3 > 0$, and change λ_1 and λ_2 into high eigenvalues to control R_B close to 1, which differentiates blob-like and plate-like lesion structures from others. Threshold parameters are $\alpha = 1$, $\beta = 0.6$, and $\gamma = 0.0444$ to balance sensitivity differentiating blob-like and plate-like lesions from background noises. To capture lesions with various sizes, we modify the multiscale value, "s", from 1 to 9 with a step size of 0.2. A few examples comparing original ROIs and responses after applying the new function are shown in Figure 4.

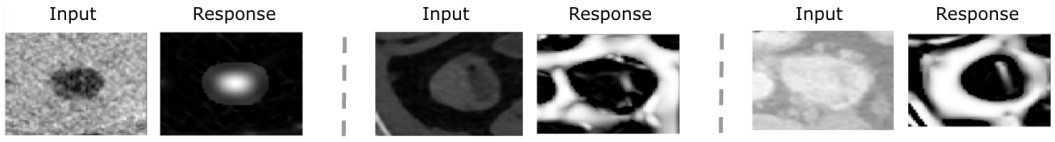

Figure 4: Sample responses after the modified Frangi filter for lesion contour detection.

C CBIR DESCRIPTORS FOR THE MOST SIMILAR LESIONS.

We used following loss function: (Gómez (2019))

$$L(a, p, n) = \frac{1}{2} max(0, m + d(a, p) - d(a, n)) \tag{2}$$

a is an anchor, p is a positive sample sharing the same lesion type with a. n is a negative sample different from a's type. m is the margin determining the stretch for separating negative and positive samples. We use cosine distance similarity measure *d* to imply how similar a retrieved candidate is to a given query:

$$d(i, j) = \frac{i \cdot j}{\|i\| \, \|j\|} \tag{3}$$

i is a query and j is a lesion candidate embedding.

D EXPERIMENTAL SETUP

Data pre-processing. We preprocess DeepLesion by cropping bounding boxes, flips, color distortions, and Gaussian blur. We keep image size as 64 x 64 and apply the Frangi filter.

Implementation Details. We train models with Pytorch (Krizhevsky et al. (2012)) on a single Nvidia DGX-A100 GPU. We implement ResNet-18 (He et al. (2016)) in SimCLR, followed by two more convolutional layers and a GeM pooling with L2-normalization layer. We performed 1000 epochs for SSL pre-training, with LARS (You et al. (2017)) optimizer of a 0.05 learning rate, a 10^{-5} weight decay, and a cosine learning rate scheduler. For CBIR task, we compute contrastive loss with margin 0.8 and cosine distance from the sigmoid classification layer. Optimized with SGD, we fine-tune the model for 50 iterations with a learning rate of 0.01, momentum of 0.9, tracker of cosine growth for the learning rate.