Gaze-Based Attention to Improve the Classification of Lung Diseases

Maksim Kholiavchenko[a], Ilya Pershin[b], Bulat Maksudov[b], Tamerlan Mustafaev[b], Yixuan YUAN[c], and Bulat Ibragimov[b, d]

[a]Rensselaer Polytechnic Institute, New York, USA
[b]Innopolis University, Innopolis city, Russia
[c]City University of Hong Kong, Hong Kong, China
[d]University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Detection of lung diseases from chest X-rays has been of great interest from the research community during the last decade. Despite the existence of large annotated public databases, computer-aided diagnostic solutions still fail on challenging rare abnormality cases. In this study, we investigated the paradigm of combining the analysis of chest X-rays and physician gaze patterns during the analysis of these X-rays to improve the computerized diagnostic accuracy. Tobii Eye Tracker 4C has been mounted to a physician workstation and his eye movements were recorded during the analysis of 400 chest X-rays in two days of work. The X-rays have been sampled from CheXpert, RSNA, and SIIM-ACR public databases labeled with 14 different pathology types. The task was formulated as a binary classification problem. A ResNet34-based neural network has been trained to map the input chest X-ray with the output physician gaze map and binary pathology label. The proposed network improved the diagnostic accuracy to 0.714 of the area under receiving operator curve (AUC) from 0.681 AUC obtained for the same ResNet34 trained to generate binary pathology labels alone. The proposed study has demonstrated the potential benefits of using gaze information in computerized diagnostic solutions.

Keywords: Eye-Tracking, Deep Learning, Classification, Segmentation

1. INTRODUCTION

Visual interpretation of medical images is the key step in modern diagnostics. During 4 years of residency in the US, a radiology trainee develops the skill to ensure comprehensive and efficient image reading.[1] Efficiency is needed to cope with the rapidly growing amount of medical images acquired per patient in modern hospitals.[2] Despite promising results shown by machine learning autodiagnosis, such systems are still far from being used in actual practice. In general, to create a robust system, it is necessary to have a dataset with very clean and correct annotations. But most modern datasets in one way or another contain noise or incorrect annotations. For example, two large public datasets were compared by visually exploring.[3] The aim of the research under consideration was to find the existing problems of such datasets and determine how accurate they are. Visual inspection showed that there are significant problems with these datasets. The analysis of radiologist's eye movements has been of growing interest since the development of eye-tracking devices.[4] Eye movements over medical images can be subdivided into fixations, i.e. focused attention on a specific location, and saccades, i.e. rapid movements between fixations.

Our paper presents an approach to improve chest x-ray classification accuracy by introducing gaze-based attention. Gaze-based attention is additional eye-tracking information that also takes part in deep neural network training. In our approach, we integrate gaze-based attention to the network in such a way that the model receives more information from the data and learns better. We trained 2 versions of a network with the same backbone architecture. The first one is an ordinary CNN model and the second one is a network with embedded gaze-based attention. We demonstrated that a model with gaze-based attention performs better in terms of accuracy. Our experiments show that gaze-based attention can be easily integrated into most modern network architectures.

Further author information: (Send correspondence to Bulat Ibragimov)
Bulat Ibragimov: E-mail: bulat@di.ku.dk

2. METHODOLOGY

2.1 Data Collection Protocol

An in-house framework mimicking a radiologist workstation was installed at our research facility. A practicing radiologist with experience in X-ray analysis was recruited to participate in the gaze analysis experiment. The radiologist was unaware of the experiment's aims except that we wanted to record his eye movements during work. We used the Tobii Eye Tracker 4C[5] as the main hardware device for eye-tracking. The tracker was set up in a quiet environment without external interference. The experiment has been divided into two days, analysing 400 unique x-rays each. Tobii Eye Tracker 4C allows for 9-point calibration which occurs every 100 images. Special minimalistic GUI has been developed to allow the radiologist to make his decision (norm or pathology) and to proceed to the next image. In this way the radiologist was able to move to the next CXR image when he was done with a given image, making the experiment easier.

During the experiment gaze coordinates, distance to the monitor, head position values were collected. The tracker was calibrated after every 100 images. After the analysis of 200 X-rays, the radiologist had a lunch break.

2.2 Model Architectures

To demonstrate the efficiency of our method, we have developed two models. Both models have the same backbone but were trained using different losses and one of the models has a decoder block. For our experiments, we used a 34-layers residual network[6] as a backbone. The model architectures are shown in Fig. 1.

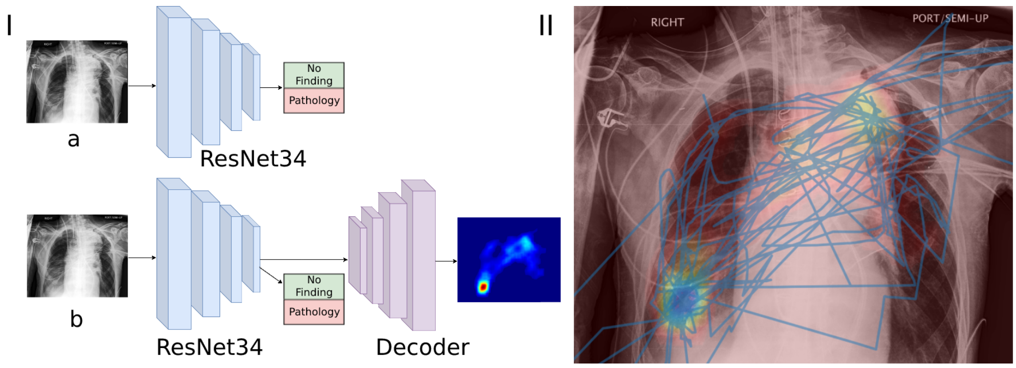

Figure 1. I.a) The first model consists of the ResNet34 model that takes chest x-ray images as inputs and predicts binary labels. I.b) The second model is similar to the first one except for the decoder part. The model takes chest x-ray images as input and predicts binary labels as well as eye-tracking heatmaps. II) Example chest x-ray image with gaze trajectory and gaze heatmap displayed.

The second model learns to predict heatmaps of eye-tracking data as well as to classify pathologies. Such architecture allows the second model to receive more information during training and also affects classification brunch due to backpropagation. We processed original images and heatmaps to make a square shape with a side of 512 pixels. The heatmaps were also binarized before training. To prevent overfitting, we used such augmentation techniques as flipping, rotation, scaling, and shearing.

2.3 Loss Functions

The first model is trained with a binary cross-entropy loss function that is widely known for its usage in classification tasks:

$$L(v, \hat{v}) = -\left(v \log \hat{v} + (1 - v) \log \left(1 - \hat{v}\right)\right), \tag{1}$$

where v is the ground truth value and $\hat{v}$ is the value predicted by our model. The second model is trained with the following loss function:

$$L(v, \hat{v}) = -\left(v \log \hat{v} + (1 - v) \log \left(1 - \hat{v}\right)\right) - \sum_{p \in P} y_p \log \left(x_p\right) - \log \frac{2 \sum_{p \in P} x_p y_p}{\sum_{p \in P} x_p^2 + \sum_{p \in P} y_p^2}, \tag{2}$$

where x is the heatmap predicted by our model, y is the corresponding ground-truth heatmap, and P is the set of pixel indexes in the heatmaps.

2.4 Training Process

We followed the same training protocol for the ResNet34 (Fig. 1a) model and for the ResNet34 model with Gaze-Based Attention (Fig. 1b). The dataset was partitioned into three folds and was trained and evaluated according to the three-fold cross-validation technique. We used Adam optimizer[7] with an initial learning rate set to 0.001 that was reduced by a factor of 0.4 each time when training reached a plateau. The batch size was set to 5 and the maximum number of epochs was set to 80. During the training of ResNet34 with Gaze-Based Attention, we froze decoder weights for the first 10 epochs, after that we unfroze decoder weights and followed the described procedure.

3. EXPERIMENTS AND RESULTS

3.1 Dataset

In this study were used 800 chest X-rays images from the databases CheXpert,[8] RSNA,[9] and SIIM-ACR[10] databases. The CheXpert database developed at Stanford University has X-rays labeled with 14 types of chest findings including 13 abnormalities and 1 no-finding label. The RSNA database was prepared by the Radiological Society of North America for the computerized detection of pneumonia. The SIIM-ACR pneumothorax database was developed and annotated by the Society for Imaging Informatics in Medicine. 10 samples were excluded from the analysis due to technical problems during the experiment. The final version of the dataset contains 790 anterior-posterior (AP) chest x-ray images and heatmaps by our eye-tracking approach. Each radiograph contains a binary label (no findings: 0 or pathology: 1).

3.2 Results and Discussion

Tab. 1 shows classification results for ResNet34 and for ResNet34 with Gaze-Based Attention models. The gaze-based attention technique helped us to improve ROC AUC by 3.3 and binary accuracy by 3.4. Such an improvement was obtained only by adding Gaze-Based Attention without changing any other hyper-parameters of our pipeline.

Table 1. Results for ResNet34 and ResNet34 with Gaze-Based Attention models.

Model	ROC AUC	Binary Accuracy
ResNet34	0.681 ± 0.019	0.666 ± 0.022
ResNet34 with Gaze-Based Attention	0.714 ± 0.016	0.700 ± 0.017

In this paper, we have validated the assumption that the radiologist's eye-tracking data contains useful information that can significantly improve classification models. In our studies, we found that while the radiologist's gaze is clinging to the pathological area of the chest x-ray images we have sufficient time to process raw eye-tracking data and create heatmaps where high intensities will be highly correlated with pathological areas or such areas that will be specific for a given pathology.

As a side effect, our model is able to predict heatmaps of chest x-ray images. This property can be useful for radiologist as assistant systems which will be able to provide an area of interest that is worth paying attention to when examining chest radiographs.

4. CONCLUSIONS

In this paper, we introduce a novel gaze-based attention method to improve the classification accuracy of lung diseases. Our approach employs eye-tracking data to be used as a source of additional information that is provided to the convolutional neural network. Our experiments show that classification accuracy can be significantly improved only by applying gaze-based attention without changing other parameters of a model. The gaze-based attention can be easily integrated into most of the modern architectures of convolutional neural networks and eye-tracking data can be collected seamlessly for a radiologist making our approach easy to integrate into existing systems for the classification of pulmonary diseases. In the future, we plan to conduct experiments with large amounts of data and merge eye-tracking data from different radiologists to create more accurate gaze-based attention maps.

ACKNOWLEDGMENTS

This research was supported by the Russian Science Foundation under Grant no.18-71-10072.

REFERENCES

[1] Bertram, R., Kaakinen, J., Bensch, F., Helle, L., Lantto, E., Niemi, P., and Lundbom, N., "Eye movements of radiologists reflect expertise in ct study interpretation: A potential tool to measure resident development," *Radiology* **281**, 805–815 (2016). doi.org/10.1148/radiol.2016151255.

[2] Bhargavan, M., Kaye, A., Forman, H. P., and Sunshine, J., "Workload of radiologists in united states in 2006-2007 and trends since 1991-1992," *Radiology* **252** (2009). doi.org/10.1148/radiol.2522081895.

[3] Oakden-Rayner, L., "Exploring large-scale public medical image datasets," *Academic radiology* **27**, 106–112 (2020). doi.org/10.1016/j.acra.2019.10.006.

[4] Alexander, R. G., Waite, S., Macknik, S. L., and Martinez-Conde, S., "What do radiologists look for? advances and limitations of perceptual learning in radiologic search," *Journal of Vision* **20**(10) (2020). doi.org/10.1167/jov.20.10.17.

[5] "Tobii Is the World Leader in Eye Tracking." https://www.tobii.com/. (Accessed: 16 August 2021).

[6] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 770–778 (2016). doi.org/10.1109/CVPR.2016.90.

[7] Kingma, D. P. and Ba, J. in [*3rd International Conference on Learning Representations, (ICLR) 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*], (2015). arxiv.org/abs/1412.6980.

[8] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., s., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D., Halabi, S., Sandberg, J., Jones, R., Larson, D., Langlotz, C., Patel, B., Lungren, M., and Ng, A., "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in [*Proceedings of the AAAI Conference on Artificial Intelligence*], **33**, 590–597 (2019). doi.org/10.1609/aaai.v33i01.3301590.

[9] "RSNA Pneumonia Detection Challenge." https://kaggle.com/c/rsna-pneumonia-detection-challenge/. (Accessed: 16 August 2021).

[10] Filice, R., Stein, A., Wu, C., Arteaga, V., Borstelmann, S., Gaddikeri, R., Galperin-Aizenberg, M., Gill, R., Hobbs, S., Jeudy, J., Lakhani, P., Laroia, A., Nayak, S., Parekh, M., Prasanna, P., Shah, P., Vummidi, D., Yaddanapudi, K., and Shih, G., "Crowdsourcing pneumothorax annotations using machine learning annotations on the nih chest x-ray dataset," *J Digit Imaging* . **33**, 490–496 (2020). doi.org/10.1007/s10278-019-00299-9.