Visualizing abnormalities in chest radiographs through salient network activations in Deep Learning

R. Sivaramakrishnan*, S. Antani, *Senior Member, IEEE*, Z. Xue, S. Candemir, S. Jaeger and G. R. Thoma

***Abstract*— This study aims to visualize salient network activations in a customized Convolutional Neural Network (CNN) based Deep Learning (DL) model, applied to the challenge of chest X-ray (CXR) screening. Computer-aided detection (CAD) software using machine learning (ML) approaches have been developed for analyzing CXRs for abnormalities with an aim to reduce delays in resource-constrained settings. However, field experts often need to know how these techniques arrive at a decision. In this study, we visualize the task-specific features and salient network activations in a customized DL model towards understanding the learned parameters, model behavior and optimizing its architecture and hyper-parameters for improved learning. The performance of the customized model is evaluated against the pre-trained DL models. It is found that the proposed model precisely localizes the abnormalities, aiding in improved abnormality screening.**

Keywords*— visualization; saliency; deep learning; machine learning; customization; activations; screening

I. INTRODUCTION

Chest X-ray (CXR) imaging diagnostics are commonly recommended for cardiopulmonary symptoms [1]. There is a significant lack of radiologists, mainly in disease-prone regions of the world, leading to an ever-growing backlog. Lack of expertise in interpreting radiology reports has been reported, especially in tuberculosis (TB) endemic regions, which is a comorbidity of HIV/AIDS, severely impairing screening efficacy [2]. Thus, current research is focused on developing cost-effective, computer-aided detection (CAD) systems based on machine learning (ML) approaches to assist radiologists in triaging and interpreting CXR images [3]. However, they are not clear about how these algorithms arrive at a decision. Visualizing the features and network activations in a model could help understanding the learned parameters and its behavior [4]. ML techniques have been previously applied to detect abnormalities in CXRs [5]–[12]. Prior works use "hand-engineered" features that demand expertise in analyzing the input variances and account for the changes in size, position, background, and orientation of the region of interest (ROI). To overcome challenges of devising high-performing hand-crafted features that capture the variance in the underlying data, Deep Learning (DL), also known as hierarchical machine learning, is used with significant success [13]. A convolutional neural network (CNN) based DL model uses a cascade of layers of nonlinear processing units for end-to-end feature extraction and classification [14]. Transfer Learning (TL) methods are commonly used to relieve problems due to data inadequacy where DL models are pre-trained on a large scale dataset like ImageNet, containing 15 million annotated images from over 22,000 classes [15]. These models produce useful features as long as the analyzed images do not deviate much from the data on which the models are trained. Biomedical images are unique to the internal body structures and have less in common with the natural images. Under these circumstances, a customized DL model can be optimized to learn task-specific features to aid in improved performance. A customized model is highly compact, flexible, has less trainable parameters and results in faster learning and convergence. The learned features and salient network activations can be visualized to understand the strategy that the model adapts to learn these task-specific features.

The goal of this study is to visualize abnormalities in CXRs through salient network activations in a customized DL model. These are used to understand the learned parameters and optimize the DL model architecture and hyper-parameters toward improved classification of normal and abnormal CXRs and localization of abnormalities. The remainder of this paper is organized as follows: Section 2 illustrates the materials and methods; Section 3 discusses the results; Section 4 gives the conclusion.

II. MATERIALS AND METHODS

A. Data collection and Preprocessing

This study uses two publicly available datasets from Montgomery County, Maryland, and Shenzhen, China, maintained by the National Library of Medicine (NLM), National Institutes of Health (NIH) [16]. Fig. 1 (a) – (e) shows some instances of abnormal and normal CXRs. Montgomery dataset contains 58 cases, tested positive for TB and 80 healthy controls. China dataset consists of 662 CXRs that include 336 cases, tested positive for TB and 326 healthy controls. Ground Truth (GT) information has been made available in the form of clinical readings, annotating the abnormal locations. The acquisition and sharing of datasets are exempted from NIH IRB review (#5357). Lung areas constituting the ROI are segmented by a method that employs anatomical atlases with non-rigid registration [17]. The ROI is resized to a 1024×1024 matrix and contrast-enhanced by applying Contrast Limited Adaptive Histogram Equalization (CLAHE) processing. Datasets are split into training (70%), validation (20%) and test (10%) and the images are augmented by translations and rotations.

R. Sivaramakrishnan is with the National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, MD 20894 USA (phone: 301-827-2383; fax: 301-402-0341; e-mail: rajaramans2@ mail.nih.gov).

Figure 1. CXRs showing: (a) hyper-lucent cystic lesions in the upper lobes, (b) right pleural effusion, (c) left pleural effusion, (d) cavitary lung lesion in the right lung and (e) normal lung.

B. Model configuration

This study evaluates the performance of a customized DL model against the pre-trained models in the task of classifying and localizing abnormalities in CXRs. We propose a sequential CNN model, having five convolutional layers and three fully connected layers. The input to the model constitutes CXRs of dimension 227×227×3. The first convolutional layer has 96 filters, each of dimension 7×7 with 2-pixel strides. The sandwich design of convolutional/Rectified Linear Unit (ReLU) nonlinearity enhances learning [18]. For the remaining convolutional layers, all the filters have a 3 × 3 receptive field. Weights are initialized from the Gaussian distribution with zero mean and standard deviation of 0.01. A local response normalization (LRN) layer follows the first and second convolutional layers that aids in generalization, motivated by a lateral inhibition process found in biological neural networks [19]. Max-pooling layers, with a pooling window of 3 × 3 and stride 2, summarize the outputs of neighboring neuronal groups in a given kernel map. The response-normalized and pooled output of the first convolutional layer is fed to the second convolutional layer, having 128 filters. The normalized and pooled output of the second convolutional layer is fed to the third convolutional layer with 256 filters. No intervening normalization and pooling layers are present between the third, fourth, and fifth convolutional layers. The fourth and fifth convolutional layers have 256 filters each. The first and second fully connected layers have 4096 neurons each, and the third fully connected layer feeds two neurons as input to the Softmax classifier. Dropout regularization with a dropout ratio of 0.5 is applied to the first and second fully connected layers. The model is trained by optimizing the multinomial logistic regression objective using stochastic gradient descent (SGD) [20] with momentum.

The customized model is optimized for hyper-parameters by a randomized grid search method [21]. Training is regularized with L2 - regularization, setting the penalty multiplier to 0.0005. Regularization helps in reducing the training error and converge to a better solution. A learning rate of 0.001 is used equally for all the layers and manually adjusted through the training process. The learning rate is divided by 10 when the validation accuracy ceased to improve and is reduced thrice before convergence. Training is stopped after 15,000 iterations (60 epochs). The customized model converges to an optimal solution due to hyper-parameter optimization, implicit regularization imposed by smaller convolutional filter sizes, greater depth, usage of L2-regularization and aggressive dropouts in the fully connected layers.

C. Usage of pre-trained models

Due to the scarcity of annotated medical imagery, TL methods are often used where a pre-trained DL model is fine-tuned to learn the current task. In this study, the last-three layers of the pre-trained models are fine-tuned for the binary classification task of interest. All the layers from the pre-trained models are extracted, except the last three layers that are replaced with a fully-connected layer, a Softmax, and a classification output layer. The fully-connected layer is set to have the size of the number of classes in the underlying data. The learning rate for the weights and biases of the fully-connected layer is increased to promote faster learning in the new layers as compared to the transferred layers. The pre-trained models are initialized to a learning rate of $1e^{-3}$ and run for 60 epochs. Since the pre-trained models have already been trained on a large-scale image dataset, we fine-tune these models and use a learning rate smaller than that used to train the models from the scratch. The models are trained on a system having Intel® Xeon® CPU E5-2640v3 2.60-GHz processor, 1 TB of hard disk space, 16 GB RAM, CUDA-enabled Nvidia GTX 1080-Ti 11GB graphical processing unit (GPU) with Windows®, Matlab® R2017a and CUDA 8.0/cuDNN 5.1 dependencies for GPU acceleration.

III. RESULTS AND DISCUSSIONS

A. Feature Visualization

The convolutional layers of the customized model output multiple channels, each corresponding to a filter applied to the input layer. The fully connected layers output channels corresponding to an abstracted version of the features learned by the earlier layers. We visualize the filters at various layers of the customized/pre-trained models as shown in Fig. 2 (a) – (l). It is observed that the customized model excels in learning task-specific features in comparison to the pre-trained models. The first convolutional layer appears to learn mostly colors and edges, indicating that the channels are color filters and edge detectors. As we progress to the third convolutional layer, we observe that the customized model learns task-specific features, including the texture of the organs with defined edges and orientations. In contrast, the pre-trained models' notion of CXRs appear to be camouflaged; they learn additional information, about the natural images on which they are trained. The third fully connected layer towards the end of the model loosely resembles the abnormal and normal classes respectively, in comparison to the pre-trained models, bearing sub-optimal resemblance to the underlying data.

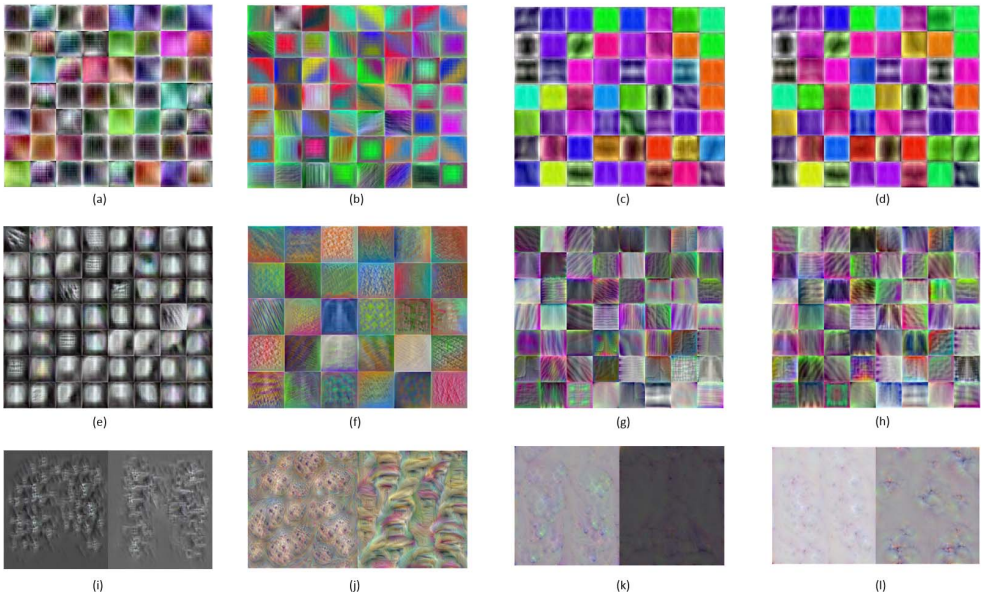

Figure 2. Visualizing the convolutional filters of the customized model, AlexNet/VGG16/VGG19 in rows. From left to right: (a) conv1, (b) conv1, (c) conv1_2, (d) conv1_2, (e)conv3, (f) conv3 (g) conv3_3, (h) conv3_4, (i) – (l) FC3.

B. Visualizing Activations

An abnormal CXR is fed into the customized/pre-trained models, and the activations of different network layers are analyzed to discover the learned features by comparing the areas of activation with the original image. The performance of the models is evaluated by investigating the activation of the channels on a set of input images. The resulting activations are compared with that of the original image, as shown in Fig. 3 (a) – (t). The channels of the first convolutional layer in the customized model are analyzed to observe the areas activating on the image and compared to the corresponding areas in the original image. All activations are scaled to the range [0, 1]. Strong positive activations are represented by white pixels and strong negative activations, by black pixels. A gray pixel does not activate as strongly on the input image. The position of a pixel in the channel activation corresponds to that in the original image. CNN learns to detect complicated features in deeper convolutional layers that build up their features by combining features from the earlier layers. The last convolutional layer, i.e., the 5th convolutional layer of the customized model and pre-trained AlexNet [19], the last convolutional layer of the 5th convolutional block of pre-trained VGG16 and pre-trained VGG19 [22] are analyzed. The channels showing strongest activation on the abnormal locations of the input image are investigated that corresponds to the 5th, 18th, 156th and 178th channels in the customized model, pre-trained AlexNet, pre-trained VGG16, and pre-trained VGG19, respectively. These channels show both positive and negative activations. However, only positive activations are investigated because of the ReLU non-linearity following the convolutional layers.

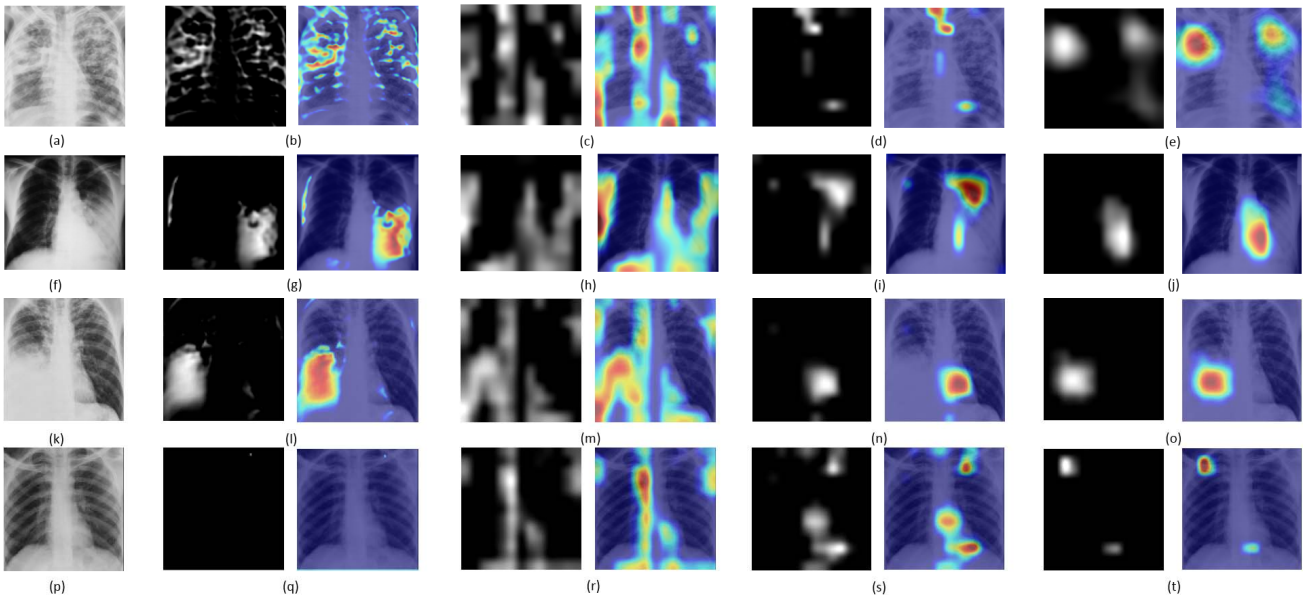

Figure 3. Visualizing the highest channel activations and heat maps in the deepest convolutional layer of the customized model, AlexNet/VGG16/VGG19 in rows. From left to right: (a) original image, (b) – (t) activations/heat maps.

The activations of the ReLU layer show the location of abnormalities. CXRs showing bilateral pulmonary TB, left pleural effusion, right pleural effusion and normal lung are input to the customized/pre-trained models, respectively. After extracting the saliency maps, the grayscale image showing the network channel activations, a pseudo color image is generated to get a clearer and more appealing representation from the perceptual aspect. A range of [0, 1] for the "jet" colormap is adapted so that the activations higher than a given threshold appear bright red, with discrete color transitions in between. The threshold is thus selected to match the range of activations and achieve the best visualization effect. The resulting heat maps are overlaid onto the original image, and the black pixels in the heat maps are made fully transparent. The more reddish the region is, the more the network activation and the more likely the area is abnormal. It is observed from the heat maps that the customized model precisely activates on the location of abnormalities, as compared to pre-trained models that exhibit sub-optimal localization behavior, increasing the false-positive rates. The customized model precisely learns task-specific features and generalizes to the data better than the pre-trained models. Learning to localize the abnormalities precisely helps the model to distinguish between normal and abnormal chest radiographs.

IV. Conclusion

The study shows that a customized CNN based DL model, unlike pre-trained models, results in the best solution for task-specific learning and localization to aid in improved screening for abnormalities in CXRs. DL models serve as triage, minimize patient loss and reduce delays in resource-constrained settings. The customized model is optimized for its architecture and hyper-parameters for improved performance and assists visualizing the learned features and layer activations toward studying its behavior. In comparison to the pre-trained models, the proposed model has fewer parameters resulting in enhanced learning, less model complexity and computation time. The proposed model can be adapted to improve the accuracy of screening for other health-related applications significantly. Next steps in our work aim to expand our analysis of customized DL models and correlate visualizations with radiology reports.

Acknowledgment

This work is supported by the Intramural Research Program of NLM, NIH and Lister Hill National Center for Biomedical Communications (LHNCBC).

References

[1] A. Bhalla, A. Goyal, R. Guleria, and A. Gupta, "Chest tuberculosis: Radiological review and imaging recommendations," *Indian J. Radiol. Imaging*, vol. 25, no. 3, p. 213, 2015.

[2] J. Melendez *et al.*, "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information.," *Sci. Rep.*, vol. 6, p. 25265, 2016.

[3] S. Jaeger *et al.*, "Automatic screening for tuberculosis in chest radiographs: a survey," *Quant. Imaging Med. Surg.*, vol. 3, no. 2, pp. 89–99, 2013.

[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2014, vol. 8689, no. 1, pp. 818–833.

[5] S. Jaeger *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 233–245, 2014.

[6] A. Chauhan, D. Chauhan, and C. Rout, "Role of gist and PHOG features in computer-aided diagnosis of tuberculosis without segmentation," *PLoS One*, vol. 9, no. 11, pp. 1–12, 2014.

[7] K. C. Santosh, S. Vajda, S. Antani, and G. R. Thoma, "Edge map analysis in chest X-rays for automatic pulmonary abnormality screening," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 9, pp. 1637–1646, 2016.

[8] M. Ding *et al.*, "Local-global classifier fusion for screening chest radiographs," in *Proc. SPIE Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*, 2017, p. 101380A.

[9] S. Katsuragawa and K. Doi, "Computer-aided diagnosis in chest radiography," *Comput. Med. Imaging Graph.*, vol. 31, no. 4–5, pp. 212–223, 2007.

[10] J. M. Carrillo de Gea, "Detection of Normality / Pathology on Chest Radiographs Using Lbp," in *Proceedings of the First International Conference on Bioinformatics, Valencia, Spain*, 2010, pp. 167–172.

[11] C. S. Venegas-Barrera and J. Manjarrez, "Visual Categorization with Bags of Keypoints," *Rev. Mex. Biodivers.*, vol. 82, no. 1, pp. 179–191, 2011.

[12] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 733–746, 2011.

[13] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8 : Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," *arXiv Prepr.*, 2017.

[14] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[15] F. Deng et al., "ImageNet: a Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[16] S. Jaeger, S. Candemir, S. Antani, J. Wang, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imaging Med. Surg.*, vol. 4, no. 6, pp. 475–477, 2014.

[17] S. Candemir *et al.*, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 577–590, 2014.

[18] W. Shang, K. Sohn, D. Almeida, and H. Lee, "Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units," in *ICML'16 Proceedings of the 33rd International Conference on International Conference on Machine Learning*, 2016, vol. 48, pp. 2217–2225.

[19] A. Krizhevsky, I. Sutskever, and H. Geoffrey, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[20] Y. LeCun, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[21] J. Bergstra and B. Yoshua, "Random Search for Hyper-Parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[22] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *Int. Conf. Learn. Represent.*, pp. 1–14, 2015.