CV - Pneumonia Detection

Project synopsis

1.0 Problem definition

Pneumonia is an inflammatory condition of the lung that primarily affects the small air sacs called alveoli, and it is caused by bacterial or viral infection. Symptoms include cough with phlegm, chest pain, difficulty breathing, and fever, and the severity of symptoms can vary. Pneumonia diagnosis is a time-consuming process that involves highly skilled professionals to analyze a chest radiograph or chest X-ray (CXR) and confirm the diagnosis with clinical history, vital signs, and laboratory tests. It helps doctors to work out the extent and placement of the infection in the lungs. Respiratory illness manifests as a neighborhood of inflated opacity on X-Ray. However, pneumonia diagnosis is complicated because increased opacity on CXRs could represent several other lung conditions, such as pulmonary edema, bleeding, volume loss, and lung cancer.

In recent years, deep-learning methods based on convolutional neural networks (CNNs) have exhibited increasing potential and efficiency in image recognition tasks, such as robotics, self-driving cars, and medical applications. For application to CXR, deep-learning models can achieve detection performance close to that of radiologists.

2.0 Literature Survey

2.1 Deep Learning Uses in Medical Imaging

2.2.1. Image Classification.

Image or exam classification was one of the first areas in which deep learning contributed majorly to medical image analysis. In exam classification, one typically has one or multiple images (an exam) as input, with a single diagnostic variable as output (e.g. disease present or not). In such a setting, every diagnostic exam is a sample and dataset sizes are typically small compared to those in computer vision (e.g., hundreds/thousands vs. millions of samples). The popularity of transfer learning for such applications is therefore not surprising.

Transfer learning is the use of pre-trained networks (typically on natural images) to work around the (perceived) requirement of large data sets for deep network training. Two transfer learning strategies were identified: (1) using a pre-trained network as a feature extractor and (2) fine-tuning a pre-trained network on medical data. The former strategy has the extra benefit of not requiring one to train a deep network at all, allowing the extracted features to be easily plugged in to existing image analysis pipelines. Both strategies are popular and have been widely applied. However, few authors perform a thorough investigation in which strategy gives the best result.

The medical imaging community initially focused on unsupervised pre-training and network architectures, like SAEs and RBMs, for classification. But recently, a clear shift towards CNNs can be observed. Authors also leverage unique attributes of medical data. Like using 3D convolutions instead of 2D. New layers have been developed, like edge-to-edge, edge- to-node, and node-to-graph layers. CNNs are the current standard techniques. Especially CNNs pre-trained on natural images indicated strong results, challenging the accuracy of human experts in some tasks. Last, authors have also shown that CNNs can be adapted to leverage the intrinsic structure of medical images.

2.2.2. Object (or lesion) Classification.

Object classification focuses on the classification of a small (previously identified) part of the medical image into two or more classes (e.g. Opacity classification in RSNA CXR). For many of these tasks, both local information on lesion appearance and global contextual information on lesion location are required for accurate classification. This combination is typically not possible in generic deep learning architectures. So several authors have used multi-stream architectures to resolve this in a multi-scale fashion. Incorporating 3D information is also often a necessity for good performance in object classification tasks in medical imaging. As images in computer vision are 2D natural images, networks developed in those scenarios do not directly leverage 3D information. Authors have used different approaches to integrate 3D effectively with custom architectures. So, almost all recent papers prefer the use of 'end-to-end' trained CNNs. An interesting approach, especially where object annotation to generate training data is expensive, is the integration of multiple instance learning (MIL) and deep learning. Xu et al. (2014) investigated the use of a MIL-framework with supervised and unsupervised feature learning approaches and also, hand-crafted features. The results showed that the performance of the MIL-framework was superior to handcrafted features, which closely approaches the performance of a fully supervised method. We expect such approaches to be popular in the future as well, as obtaining high-quality annotated medical data is challenging.

Overall, object classification sees less use of pre-trained networks compared to exam classifications, mostly because of the need for incorporation of contextual or three-dimensional information. Several authors have found innovative solutions to add this information to deep networks with good results, and we expect deep learning to become even more prominent for this task soon.

2.2.3. Object or lesion detection.

Anatomical object localization (in space or time), such as organs or landmarks, has been an important preprocessing step in segmentation tasks or in the clinical workflow for therapy planning and intervention. And localization through 2D image classification with CNNs seems to be the most popular strategy to identify organs, regions and landmarks, with good results.

In fact, the detection of objects of interest or lesions in images is a key part of diagnosis and is one of the most labor-intensive tasks for clinicians. It comprises localization and identification of small lesions in the full image space. There has been a long research tradition in computer-aided detection systems that can detect lesions automatically, improving the detection accuracy or decreasing the reading time of human experts.

Most of the published deep learning object detection systems still use CNNs to perform pixel (or Voxel) classification, after which some form of post processing is applied to obtain object candidates. As the classification task performed at each pixel is essentially object classification, CNN architecture and methodology are like those in of Classification. The incorporation of contextual or 3D information is handled using multi-stream CNNs. Last, as the annotation burden to generate training data can be similarly significant compared to object classification, weakly supervised deep learning has been explored by Hwang and Kim (2016), who adopted such a strategy for the detection of nodules in chest radiographs and lesions in mammography.

But there are some aspects which are significantly different between object detection and object classification. One key point is that because every pixel is classified, typically the class balance is skewed severely towards the non-object class in a training setting. To add insult to injury, most of the non-object samples are easy to discriminate, preventing the deep learning method from focusing on the challenging samples. van Grinsven et al. (2016) proposed a selective data sampling in which wrongly classified samples were fed back to the network more often to focus on challenging areas in retinal images. Last, as classifying each pixel in a sliding window fashion results in orders of magnitude of redundant calculation, fCNNs are important aspects of an object detection pipeline as well.

Challenges in meaningful application of deep learning algorithms in object detection are thus mostly similar to those in object classification. Only a few papers directly address issues specific to object detection- like class imbalance/hard-negative mining or efficient pixel/voxel-wise processing of images. We expect that more emphasis will be given to those areas in the near future, for example, in the application of multi-stream networks in a fully convolutional fashion.

2.2.4. Segmentation.

The segmentation of organs and other substructures in medical images allows quantitative analysis of clinical parameters related to volume and shape, as, for example, in cardiac or brain analysis. Furthermore, it is often an important first step in computer-aided detection pipelines. Segmentation is defined as identifying the set of voxels which make up either the contour or the interior of the object (s) of interest. Segmentation is the most common subject of papers applying deep learning to medical imaging, and, as such, has also seen the widest variety in methodology, including the development of unique CNN-based segmentation architectures and the wider application of RNNs.

The most well-known, in medical image analysis, of these novel CNN architectures is U-net, published by Ronneberger et al. (2015). The two main architectural novelties in U-net are the combination of an equal amount of up-sampling and down-sampling layers. Although learned up sampling layers have been proposed before, U-net combines them with skip connections between opposing convolution and deconvolution layers. This which concatenates features from the contracting and expanding paths. From a training perspective, this means that an entire set of images/scans can be processed by U-net in one forward pass, resulting in a segmentation map directly. This allows U-net to consider the full context of the image, which can be an advantage in contrast to patch-based CNNs.

Although these specific segmentation architectures offered compelling advantages, many authors have also obtained excellent segmentation results with patch-trained neural networks. One of the earliest papers covering medical image segmentation with deep learning algorithms used such a strategy and was published by Ciresan et al. (2012). They applied pixel-wise segmentation of membranes in electron microscopy imagery in a sliding window fashion. Most recent papers now use fCNN (fully connected Convolutional Neural Networks) over sliding- window-based classification to reduce redundant computation.

Summarizing, segmentation in medical imaging has seen a huge influx of deep learning related methods. Custom architectures have been created to target the segmentation task directly. These have given promising results, rivaling and often improving over results obtained with fCNNs.

2.2.5. Lesion segmentation.

Segmentation of lesions combines the challenges of object detection and organ and substructure segmentation in the application of deep learning algorithms. Global and local context are typically needed to perform accurate segmentation, such that multi-stream networks with different scales or non-uniformly sampled patches are used. In lesion segmentation, we have also seen the application of U-net and similar architectures to leverage both this global and local context. One other challenge that lesion segmentation shares with object detection is class imbalance, as most voxels/pixels in an image are from the non-diseased class. Some papers combat this by adapting the loss function: Brosch et al. (2016) defined it to be a weighted combi- nation of the sensitivity and the specificity, with a larger weight for the specificity to make it less sensitive to the data imbalance. Others balance the data set by performing data augmentation on positive samples.

Thus, lesion segmentation sees a mixture of approaches used in object detection and organ segmentation. Developments in these two areas will most likely naturally propagate to lesion segmentation as the existing challenges are also mostly similar.

2.2 Overview of deep learning methods

Learning algorithms.

Machine learning methods are divided into supervised and unsupervised learning algorithms. Unsupervised learning algorithms process data without labels and are trained to find patterns, such as latent subspaces. But in supervised learning, a model is presented with a dataset of input features x and label y pairs, where y typically represents an instance of a fixed set of classes. With regression tasks, y can also be a vector with continuous values. Supervised training typically amounts to finding model parameters Θ that best predict the data based on a loss function L (y, y_hat). Here, y_hat denotes the output of the model obtained by feeding a data point x to the function f (x; Θ) that represents the model.

Neural Networks.

Neural networks are a type of learning algorithm that forms the basis of most deep learning methods. A neural network consists of neurons or units with some activation and parameters W and B, where W is a set of weights and B a set of biases. The activation represents a linear combination of the input x to the neuron and the parameters, followed by an element-wise non- linearity, referred to as a transfer function.

Typical transfer functions for traditional neural networks are the sigmoid and hyperbolic tangent function. The multi-layered perceptrons (MLP), the most well- known of the traditional neural networks, have several layers of such linear + non-linear transformations. Layers in between the input and output are 'hidden' layers. When a neural network contains multiple hidden layers, it is considered a 'deep' neural network, hence the term 'deep learning'. At the final layer of the network, the activations are mapped to a distribution over classes through a Softmax function.

Maximum likelihood with stochastic gradient descent is currently the most popular method to fit parameters Θ to a dataset. In stochastic gradient descent, a small subset of the data, a mini-batch, is used for each gradient update instead of the full data set. Optimizing maximum likelihood in practice amounts to minimizing the negative log-likelihood. This results in the binary cross-entropy loss for two-class problems and the categorical cross-entropy for multi-class tasks. A downside of this approach is that it rarely optimizes the quantity we are interested in- such as the area under the receiver- operating characteristic (ROC) curve or common evaluation measures for segmentation, such as the Dice coefficient.

For a long time, deep neural networks (DNN) were considered hard to train efficiently. They only gained popularity in 2006 when it was shown that training DNNs layer-by-layer in an unsupervised manner (pre-training), followed by supervised fine-tuning of the stacked network, could result in good performance. Two popular architectures trained in such a way are stacked auto-encoders (SAEs) and deep belief networks (DBNs). However, these techniques are rather complex and require a significant amount of engineering to generate satisfactory results.

Currently, the most popular models are trained end-to-end in a supervised fashion, hugely simplifying the training process. The most popular architectures are convolutional neural networks (CNNs) and recurrent neural networks (RNNs). CNNs are currently the most widely used in (medical) image analysis, although RNNs are gaining popularity.

Convolutional Neural Networks (CNNs)

There are two key differences between MLPs and CNNs. First, in CNNs, convolution operations on images makes sure that the model does not need to learn separate detectors for the same object occurring at different positions in an image, making the network equivariant with respect to translations of the input (translational invariance). It also drastically reduces the number of parameters that need to be learned.

At each layer, the input image is convolved with a set of K kernels and added biases, each generating a new feature map. These features are subjected to an element-wise non-linear transform and the same process is repeated for every convolutional layer.

The second key difference between CNNs and MLPs is the typical incorporation of pooling layers in CNNs, where pixel values of neighborhoods are aggregated using a permutation invariant function, typically the max or mean operation. This induces a certain amount of translation invariance and again reduces the number of parameters in the network. At the end of the convolutional stream of the network, fully connected layers (i.e. regular neural network layers) are usually added, where weights are no longer shared. Similar to MLPs, a distribution over classes is generated by feeding the activations in the final layer through a Softmax function and the network is trained using maximum likelihood.

Deep CNN Architectures.

Given the prevalence of CNNs in medical image analysis, we elaborate on the most common architectures and architectural differences among the widely used models.

LeNet (LeCun et al.,1998) and AlexNet (Krizhevsky et al. 2012), introduced over a decade later, were very similar models. Both networks were relatively shallow, comprising two and five convolutional layers, respectively, and employed kernels with large receptive fields in layers close to the input and smaller kernels closer to the output. AlexNet, though, incorporated rectified linear units instead of the hyperbolic tangent as an activation function.

After 2012, the exploration of novel architectures took off. By stacking smaller kernels, instead of using a single layer of kernels with a large receptive field, a similar function can be represented with fewer parameters. These deeper architectures have a lower memory footprint during inference, which enables their deployment in mobile computing devices such as smartphones. A 19-layer model, employed small, fixed size kernels in each layer, often referred to as VGG19 won the ImageNet challenge of 2014.

On top of the deeper networks, more complex building blocks have been introduced that improve the efficiency of the training procedure and again reduce the number of parameters. GoogLeNet is a 22-layer network, also referred to as Inception, which makes use of inception blocks- a module that replaces the mapping (see the equation below) in convolution networks with a set of convolutions of different sizes.

$$\mathbf{X}_k^l = \sigma(\mathbf{W}_k^{l-1} * \mathbf{X}^{l-1} + b_k^{l-1}).$$

Similar to the stacking of small kernels, this allows a similar function to be represented with fewer parameters. The ResNet architecture won the ImageNet challenge in 2015 and comprised ResNet-blocks. Rather than learning a function, the residual block only learns the residual and is thereby pre-conditioned towards learning mappings in each layer that are close to the identity function. This way, even deeper models can be trained effectively.

Since 2014, the performance on the ImageNet bench- mark has saturated, and it is difficult to assess whether the small increases in performance can really be attributed to 'better' and more sophisticated architectures. The advantage of the lower memory footprint these models provide is typically not as important for medical applications. Consequently, AlexNet or other simple models such as VGG are still popular for medical data, though recent landmark studies all use a version of GoogleNet called Inception v3. Whether this is because of a superior architecture or simply because the model is a default choice in popular software packages is again difficult to assess.

Multi-stream architectures.

The default CNN architecture can easily accommodate multiple sources of information or representations of the input, as channels presented to the input layer. This idea can be taken further and channels can be merged at any point in the network. Under the intuition that different tasks require different ways of fusion, multi-stream architectures are being explored. These models, also referred to as dual pathway architectures, have two main applications at the time of writing: (1) multi-scale image analysis and (2) 2.5D classification; both relevant for medical image processing tasks.

For the detection of abnormalities, context is often an important cue. The most straightforward way to increase context is to feed larger patches to the network, but this can significantly increase the number of parameters and memory requirements of a network. Consequently, architectures have been investigated where context is added in a down-scaled representation besides high resolution local information. Several medical applications have also successfully used this concept.

The challenge of applying deep learning techniques to the medical domain often lies in adapting existing architectures to, for instance, different input formats, such as three-dimensional data. In early applications of CNNs to such volumetric data, full 3D convolutions and the resulting large number of parameters were circumvented by dividing the Volume of Interest (VOI) into slices which are fed as different streams to a network. Similarly, the network can be fed with multiple angled patches from the 3D-space in a multi-stream fashion, which has been applied by various authors in medical imaging. These approaches are also referred to as 2.5D classification.

Segmentation Architectures

Segmentation is a common task in both natural and medical image analysis and to tackle this, CNNs can be used to classify each pixel in the image individually, by presenting it with patches extracted around the particular pixel. A drawback of this naïve ‘sliding- window’ approach is that input patches from neighboring pixels have huge overlap and the same convolutions are computed many times. Fortunately, the convolution and dot product are both linear operators and thus inner products can be written as convolutions and vice versa. By rewriting the fully connected layers as convolutions, the CNN can take input images larger than it was trained on and produce a likelihood map, rather than an out- put for a single pixel. The resulting ‘fully convolutional network’ (fCNN) can then be applied to an entire input image or volume efficiently.

However, because of pooling layers, this may result in output with a far lower resolution than the input. ‘Shift-and-stitch’ is one of several methods proposed to prevent this decrease in resolution. The fCNN is applied to shifted versions of the input image. By stitching the result together, one obtains a full resolution version of the final output, minus the pixels lost because of the ‘valid’ convolutions.

Ronneberger et al. (2015) took the idea of the fCNN one step further and proposed the U-net architecture, comprising a ‘regular’ fCNN followed by an up-sampling part where ’up’-convolutions are used to in- crease the image size, coined contractive and expansive paths. The authors then combined it with so-called skip-connections to connect opposing contracting and expanding convolutional layers directly. Milletari et al. (2016b) proposed an extension to the U-Net layout that incorporates ResNet-like residual blocks and a Dice loss layer, rather than the conventional cross-entropy, that directly minimizes this commonly used segmentation error measure.

2.3 The context of Pneumonia Detection

Roth et al. demonstrated the power of deep convolutional neural networks (CNN) to detect the lymph node in clinical diagnostic tasks and got drastic results even in the presence of low contrast surrounding structures obtained from computer tomography. In another study, Shin et al. addressed the problems of thoraco-abdominal lymph detection and interstitial lung disease classification using deep CNN. They developed different CNN architectures and got promising results with 85 percent sensitivity at three false positives per patient. Ronneburger et al. developed a CNN approach with the use of data augmentation. They suggested that even trained on small samples of image data obtained from transmitted light microscopy; the developed model could capture high accuracy. Jamaludin et al. applied CNN architecture to analyze the data obtained from spinal lumber magnetic resonance imaging (MRI). They developed an efficient CNN model to generate radiological grading of spinal lumber MRIs.

All these studies have performed well on radiological data, except that the size of the data was restricted to a few hundred samples of patients. Therefore, a detailed study is required to use the power of deep learning over thousands of samples of patients to achieve accurate and reliable predictions.

Kallianos et al. presented a state of art review stating the importance of artificial intelligence in chest X-ray image classification and analysis. Wang et al. prepared a new database ChestX-ray8 with 108,948 front view X-ray images of 32,717 unique patients. Each of the X-ray images could have multiple labels. They used deep convolutional neural networks to validate the results on this data and obtained promising results. They mentioned that the chestX-ray8 database can be extended by including more disease classes and would be useful for other research studies.

Rajpurkar et al. developed CheXNet10 by using the ChestX-ray14 dataset11, which contains 112,120 frontal-view chest x-ray images individually labeled with 14 different pathologies, including pneumonia. Tey estimated and compared the performance of the model and four radiologists, revealing that the model exceeded the average performance of the radiologists on the pneumonia detection task. Hwang et al. applied a deep-learning method for chest radiograph diagnosis in the emergency department, with their results showing that the diagnostic performance of radiology residents using CXR readings improved after radiographs were reinterpreted through the deep-learning algorithm output.

Irvin et al. stated that a large labeled dataset is the key to success in prediction and classification tasks. They presented a huge dataset that comprises 224,316 chest radiographic images of 65,240 patients. They named this dataset as CheXpert. Then they used convolutional neural networks to assign labels to them based on the probability assigned by the model. Model used frontal and lateral radiographs to output the probabilities of each observation.

Further, they released the dataset as a benchmark dataset. Besides the availability of a large dataset, it is highly desirable that every object in the image should be detected carefully and segmentation of each instance should be done precisely. Therefore, a different approach is required to handle both instance segmentation and object detection.

Such powerful methods are faster region-based CNN (F-RCNN) and FCN (Fully Convolutional Network). F-RCNN can be extended with an additional branch for segmentation mask prediction on each region of interest, along with existing branches for the classification task. This extended network is called Mask R-CNN, and it is better than F-RCNN in terms of efficiency and accuracy. Kaiming He et al. presented Mask R-CNN approach for object instance segmentation. They compared their results with the best models from COCO 2016. Luc et al. extended their approach by introducing an instance level segmentation by predicting convolutional features.

3.0 Dataset

The labeled dataset of the chest X-Ray images and patients' metadata was publicly provided for a Kaggle challenge by the US National Institutes of Health Clinical Center. The database comprises frontal-view X-ray images from 26684 unique patients. Each image is labeled with one of three different classes from the associated radiological reports: "Normal"," No Lung Opacity / Not Normal"," Lung Opacity".

A word on Opacity

Tissues with sparse material, such as lungs which are full of air, do not absorb the X-rays and appear black in the image. Dense tissues such as bones absorb the X-rays and appear white in the image. When someone has pneumonia, the air in the lungs is replaced by other material, i.e. fluids, bacteria, immune system cells, etc. The lung opacities refer to the areas that preferentially attenuate the x-ray beam and therefore appear opaquer on CXR than they should. In the dataset, the 'Lung Opacity' class has images of the presence of white fuzzy clouds in the lungs- associated with pneumonia. These are labeled with bounding boxes. Any patient can have multiple boxes if more than one opacity region is detected.

data is shared in the below drive:

https://drive.google.com/drive/folders/1GYAe8hZB8Si5YSW0akNXBpsELRicE4Hp

4.0 Tentative list of algorithms.

We will use these algorithms with various backbone networks like ResNet or CoCo.

- Faster-RCNN
- YOLO3
- U-Net
- Mask RCNN

5.0 References

1. Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *RadioGraphics* 2017; 37:2113–2131 [Crossref] [Medline] [Google Scholar]
2. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015; 115:211–252 [Crossref] [Google Scholar]
3. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, eds. *Advances in neural information processing systems 25 (NIPS 2012).* San Diego, CA: Neural Information Processing Systems Foundation, 2012 [Google Scholar]
4. Prevedello LM, Halabi SS, Shih G, et al. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. *Radiol Artific Intell* 2019; 1:e180031 [Crossref] [Google Scholar]
5. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *IEEE Proceedings: 30th IEEE Conference on Computer Vision and Pattern Recognition.* Los Alamitos, CA: IEEE, 2017:3462–3471 [Google Scholar]
6. Shih G, Wu CC, Halabi SS, et al. Augmenting the National Institutes of Health chest radiograph dataset with expert annotations of possible pneumonia. *Radiol Artif Intell* 2019; 1:e180041 [Crossref] [Google Scholar]
7. RSNA Pneumonia Detection Challenge: overview. Kaggle website. www.kaggle.com/c/rsna-pneumonia-detection-challenge. Accessed July 17, 2019 [Google Scholar]
8. Halabi SS, Prevedello LM, Kalpathy-Cramer J, et al. The RSNA pediatric bone age machine learning challenge. *Radiology* 2019; 290:498–503 [Crossref] [Medline] [Google Scholar]
9. Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks. In: *2017 IEEE International Conference on Computer Vision (ICCV).* Piscataway, NJ: IEEE, 2017:764–773 [Google Scholar]

10. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. arXiv website. arxiv.org/abs/1602.07261. Published February 23, 2016. Accessed May 18, 2018 [Google Scholar]
11. Chollet F. Xception: deep learning with depthwise separable convolutions. arXiv website. arxiv.org/abs/1610.02357. Published October 7, 2016. Accessed May 18, 2018 [Google Scholar]
12. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. arXiv website. arxiv.org/abs/1608.06993. Published August 24, 2016. Accessed May 18, 2018 [Google Scholar]
13. ImageNet website. ImageNet. www.image-net.org/. Accessed February 20, 2019 [Google Scholar]
14. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: *2017 IEEE International Conference on Computer Vision.* Los Alamitos, CA: IEEE, 2017:2999–3007 [Google Scholar]
15. Dai J, Li Y, He K, Sun J. R-FCN: object detection via region-based fully convolutional networks. arXiv website. arxiv.org/abs/1605.06409v2. Published May 20, 2016. Accessed February 22, 2019 [Google Scholar]
16. Hu H, Gu J, Zhang Z, Dai J, Wei Y. Relation networks for object detection. arXiv website. arxiv.org/abs/1711.11575v2. Published November 30, 2017. Accessed February 22, 2019 [Google Scholar]
17. Code for 1st place solution in Kaggle RSNA Pneumonia Detection Challenge. GitHub website. github.com/i-pan/kaggle-rsna18. Accessed July 17, 2019 [Google Scholar]
18. Ng A. Convolutional neural networks. Coursera website. www.coursera.org/learn/convolutional-neural-networks. Accessed December 27, 2018 [Google Scholar]
19. Howard J. Cutting edge deep learning for coders: part 2. Onwards fast ai website. course18.fast.ai/part2.html. Accessed January 26, 2019 [Google Scholar]
20. Gaiser H. Keras implementation of RetinaNet object detection: keras-retinanet. github.com/fizyr/keras-retinanet. GitHub website. Accessed January 19, 2019 [Google Scholar]
21. 3rd Place solution for RSNA Pneumonia Detection Challenge. GitHub website. github.com/pm-cheng/rsna-pneumonia. Accessed March 2019 [Google Scholar]
22. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition.* Los Alamitos, CA: IEEE, 2009:248–255 [Google Scholar]
23. Pan I, Agarwal S, Merck D. Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks. *J Digit Imaging* 2019 Mar 5 [Epub ahead of print] [Google Scholar]
24. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Med* 2018; 15:e1002683 [Crossref] [Medline] [Google Scholar]
25. Chang K, Balachandar N, Lam C, et al. Distributed deep learning networks among institutions for medical imaging. *J Am Med Inform Assoc* 2018; 25:945–954 [Crossref] [Medline] [Google Scholar]
26. RSNA Pneumonia Detection Challenge: private leaderboard. Kaggle website. www.kaggle.com/c/rsna-pneumonia-detection-challenge/leaderboard. Accessed July 17, 2019 [Google Scholar]
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arxiv.org/abs/1512.03385. arXiv website. Published December 10, 2015. Accessed April 14, 2019 [Google Scholar]
Address correspondence to P. M. Cheng (phillip.cheng@med.usc.edu).