Prediction of Locations in Medical Images Using Orthogonal Neural Networks

Jong Soo Kim (jongkim57@hanyang.ac.kr)
Hanyang University
Yongil Cho
Hanyang University
Tae Ho Lim
Hanyang University

Research Article

Keywords: deep learning, glottis, localization, orthogonal neural network, pneumothorax

Posted Date: June 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-649890/v1

Abstract

An orthogonal neural network (ONN), a new deep-learning structure for medical image localization, is developed and presented in this paper. This method is simple, efficient, and completely different from a convolution neural network (CNN). The diagnostic performance of ONN for detecting the location of pneumothorax in chest X-rays was assessed and compared to that of CNN. An area under the receiver operating characteristic (ROC) curve (AUC) of 0.870, an accuracy of 85.3%, a sensitivity of 75.0%, and a specificity of 86.5% were achieved; the ONN outperformed the CNN. The diagnostic performance of the ONN with a sigmoid activation function for all the nodes obviously outperformed the ONN with the rectified linear unit (RELU) activation function for all the nodes other than the output nodes. In addition, by applying ONN and CNN to predict the location of the glottis in laryngeal images, we achieved accurate and adjacent prediction rates of 70.5% and 20.5%, respectively, with the ONN. The prediction accuracy of the ONN was compared favorably with that of the CNN. Compared to a CNN, an ONN required only approximately 10% of the computations using a CNN trained on images with an input resolution of 256 × 256 pixels. A fully-connected small artificial neural network (ANN), selected by comparing the test results of several dozens of small ANN models, achieved the best location prediction performance on medical images. This study demonstrated that an ONN can be used as a quick selection criterion to compare ANN models for image localization since an ONN performed well compared decently with the selected ANN model.

Introduction

Deep-learning technology has recently emerged as an alternative to solving difficult scientific and technical problems, including those in the medical field. Promising results have been achieved in medical image analysis including in the detection of diabetic retinopathy in fundus photographs, the detection of metastases in pathological images, and the classification of skin cancer from skin radiographs[1–3]. Thus, it can provide a fast response for patients with a suspected disease and can overcome the deterioration of human vision diagnostic performance in general, especially if the image quality is degraded due to technical issues.

The back-propagation method[4–6] has been the most popular training method for deep learning to date. In addition, convolution neural networks[7–8] (CNNs) have been a common currently used deep-learning structure for image recognition, performing especially well in classification and object detection tasks. For the back-propagation method, the gradient descent has been used to determine each weight factor in an artificial neural network (ANN) by calculating the delta value using all or a part of the training data repeatedly according to a given learning rate. Therefore, the rectified linear unit (RELU) function has been widely used as the activation function of nodes in ANN for conveniently calculating the output gradient of a given node.

Distinct from the back-propagation method, a novel deep-learning algorithm, namely, the Kim-Monte Carlo algorithm, was recently developed by J. S. Kim[9]. This algorithm imitates the biological mechanism of

animals adapting to a given environment according to the evolutionary principle of the survival of the fittest[9]. The algorithm, which is a simple training process for ANNs, does not need to calculate the output gradient of a given node in ANN during the training session as the back-propagation method does[4–6]. Therefore, it becomes not more convenient to use the RELU activation function than a sigmoid function within an ANN node.

The Kim-Monte Carlo algorithm has been applied to train ANNs to predict the location of the glottis in laryngeal images, to detect and classify intracranial haemorrhage on CT images, to detect pneumoperitoneum in abdominal radiograph images, and to detect the location of pneumothorax in chest X-rays[9–12]. Cho *et al.*[12] found that the diagnostic performance of CNN was lower than that of a small ANN for detecting the location of pneumothorax in chest X-rays, although the CNN required approximately 10 times of the computations using a small ANN trained on input X-ray images. The small ANN was selected by comparing the test results of several dozens of fully-connected small ANN models. They[12] indicated that CNNs may not perform well for image localization other than classification or object detection, and also found that the CNN with a sigmoid activation function for fully-connected hidden nodes outperformed the CNN with the RELU activation function.

In this study, an orthogonal neural network (ONN), which is simple, efficient, and completely different from a CNN, is proposed as a new deep-learning structure for localization on medical images. The performances of ONNs for detecting the location of pneumothorax in chest X-rays[12] and for predicting the location of the glottis in laryngeal images[9] were evaluated. In addition, the performances of ONNs with a sigmoid and RELU activations were assessed and compared. The performance of CNN for predicting the glottic location in laryngeal images[9] was also investigated with the same dataset for comparison.

Methods

Orthogonal neural network

Just as a CNN consists of a convolution part and a fully-connected part, an ONN is composed of an orthogonal part and a fully-connected part. Figure 1 shows a detailed schematic representation of the ONN architecture. An input image of width n and height m provides input values to the $n \times m$ input nodes of the orthogonal part, which includes vertical layers and horizontal layers, as shown in Fig. 1.

An input node is connected only to a vertical node with the same row number. If there are 2 vertical layers, the input nodes in the first half of a given row are connected to the node in the first vertical layer, and the input nodes in the second half of the given row are connected to the node in the second vertical layer. The weight factors for the connections of the input nodes in each column to a vertical layer are set to be the same. Each vertical node has its own bias value. Therefore, the vertical layers contain n weight factors and $m \times 2$ bias values if the number of vertical layers is 2 (see Fig. 1). Likewise, an input node is connected only to a horizontal node with the same column number. If there are 2 horizontal layers, the

input nodes in the first half of a given column are connected to the node in the first horizontal layer, and the input nodes in the second half of the given column are connected to the node in the second horizontal layer. The weight factors for the connections of the input nodes in each row to a horizontal layer are set to be the same. Each horizontal node has its own bias value. Therefore, the horizontal layers contain m weight factors and $n \times 2$ bias values if the number of horizontal layers is 2 (see Fig. 1).

According to Fig. 1, the orthogonal part reflects the spatial location information of the input image and produces $m \times 2 + n \times 2$ outputs for 2 vertical layers and 2 horizontal layers, providing the input values for the fully-connected part. An ONN requires significantly fewer weight factors and computing resources than a fully-connected ANN for training on images with a given input resolution. Additionally, an ONN contains a similar number of weight factors as a CNN and requires much fewer computing resources than a CNN for training on images with a given input resolution.

Datasets

The institutional review board (IRB) of Hanyang University Seoul Hospital (Seoul, Republic of Korea) approved this study and confirmed that all methods in this study were performed in accordance with the Good Clinical Practice guidelines with the need for informed consent waived (IRB No. HYUH 2021-03-024 for chest X-ray images; IRB No. HYUH 2018-08-018-002 for laryngeal images).

Cho *et al.*[12] randomly selected 1,000 chest X-ray images with pneumothorax from NIH (National Institutes of Health, Maryland, USA) Chest X-rays dataset (available at https://nihcc.app.box.com/v/ChestXray-NIHCC) and randomly divided into training (80%) and test (20%) sets for ANNs and CNNs to detect the location of pneumothorax. For each chest X-ray image, the areas with pneumothorax were marked in white and the rest were marked in black (Fig. 2). Each pneumothorax-marked image was divided into 49 boxes for pneumothorax localization (see Fig. 2b). The target values of the boxes in the pneumothorax-marked image containing one or more white pixels were set to 1, and the target values of the other boxes were set to 0[12]. ONNs were applied to the same dataset for comparison with the pneumothorax detection results of Cho *et al.*[12] for each of the boxes in the chest X-ray images.

Kim *et al.*[9] acquired 1,200 laryngeal images during intubation in an emergency room and randomly divided them into training (1,000) and test (200) sets for ANNs to predict the glottic location. For each laryngeal image, the location of the glottis was marked in white, and the rest were marked in black (Fig. 3). Each glottis-marked image was divided into 49 boxes for glottis localization (see Fig. 3b). The target value of the box in the glottis-marked image with the maximum average pixel value (Fig. 3b, red box) was set to 1, and the target values of the other 48 boxes were set to 0[9]. The same dataset was also applied to CNNs as well as ONNs for comparison with the prediction accuracy findings of Kim *et al.*[9] for the glottic location in laryngeal images.

Training process

For an input resolution of 256 × 256 pixels, each black-and-white pixel of the original images was divided by the maximum value of 255 to obtain a value between zero and 1.0 to use as the input value of an input node of the ONN and CNN. The number of hidden layers of the fully-connected part of the ONN and CNN was set to 1 or 2. The number of hidden nodes was set to 49 or 49–49. The number of output nodes in the fully-connected part was set to 49 for the pneumothorax and glottis localizations (Figs. 2b and 3b, respectively). Figure 4 shows a detailed schematic representation of the actual architecture of the CNN for predicting the glottic location in laryngeal images. According to Fig. 4, the convolution part contains 84 individual 3 × 3 convolution filters (= 4 + 16 + 64) with 3 × 3 × 84 weight factors and 84 bias values.

The Kim-Monte Carlo algorithm[9], a simple ANN training process that differs from the back-propagation method[4–6], was applied to train the ONN and CNN. The algorithm applied a random optimization process based on a Monte Carlo simulation during the training session to determine a massive number of unknown weight factors and bias values of the ONN or CNN as the variables that minimize the average training error for a given training dataset. The initial weight factors and bias values were randomly chosen within ranges of -0.2 to + 0.2 and 0 to + 0.2, respectively[9]. The algorithm consisted of (a) randomly selecting the weight factors and bias values according to a given variable selection ratio and adjusting their values by small random amounts within the range of -0.1 to + 0.1[9], where it changes the bias value to be positive when the value is changed to be negative, (b) accepting or rejecting the adjustments depending on whether the new values decreased the average training error for all training data, and (c) repeating the above two steps[9]. The training session ended after 10 training cycle iterations, during which the variable selection ratio of the training cycle was steadily decreased from 15–1.5%[9] of the total number of variables in the ONN or CNN. During a training cycle, the total sum of the variable selection ratios was set to 900%, and 30 attempts were made to adjust the values of the selected variables by small random amounts[10,11]. After the training session was completed, the test set was applied to the ONN and CNN to obtain the test results and evaluate the diagnostic performance of the deep-learning model.

Implementation

For the activation function for nodes in the ONN or CNN, a sigmoid or RELU function was applied.

$$y_j = x_j, \quad \text{if } x_j > 0$$
$$y_j = 0, \quad \text{if } x_j \leq 0 \qquad (1)$$

The RELU activation function was implemented using Equation (1) and the sigmoid activation function was implemented by the following formulas[12]:

$$y_j = \frac{2}{1 + e^{-x_j}} - 1 \qquad (2)$$

$$y_j = \frac{1}{1 + e^{-x_j}} \qquad (3)$$

$$x_j = \sum_i w_{ij} y_i + b_j \qquad (4)$$

where Equation (4) (using one of Equations (1), (2), and (3)) denotes the summation over all nodes in the previous layer; i.e., each node *j* in a given layer receives an input from a node *i* in the previous layer; w_{ij} indicates the weight factor between nodes *i* and *j*; and b_j indicates the bias value of a node *j*. The sigmoid activation functions in Equations (2) and (3) were applied to the intermediate nodes and output nodes, respectively[9].

Details about the hardware and software infrastructure used to implement the ONNs and CNNs are described below. The software for ONN and CNN, applying the novel deep-learning algorithm, a simple training process, was developed and programmed fully in-house by J. S. Kim using Microsoft Visual C++. Regular PCs without graphics processing unit (GPU) computing were used[11,12]. The operating system used was Microsoft Windows 10 Professional (64-bit), and the CPU was an Intel i5–7400 processor. The main memory size was 16 GB. The time taken to train an ONN was approximately 5 hours for the training set of 800 chest X-ray images with an input resolution of 256 × 256 pixels. It took approximately one second to predict the entire test set of 200 X-ray images. In addition, the time taken to train a CNN was approximately 60 hours for the training set of 1,000 laryngeal images with an input resolution of 256 × 256 pixels. It took a few seconds to predict the entire test set of 200 laryngeal images.

Statistical analysis

We performed a receiver operating characteristic (ROC) curve analysis to determine the diagnostic performance of the ONNs for detecting the location of pneumothorax in chest X-rays. Statistical analyses were conducted using R software [version 4.0.1] (R: A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2021, http://www.R-project.org).

Results

Detecting the pneumothorax location

Table 1 shows the pneumothorax detection results for each of the boxes in the chest X-ray images in the test set by ANNs[12], CNNs[12], and ONNs. In Table 1, AUC denotes the area under the ROC curve, PPV

indicates the positive predictive values, and NPV indicates the negative predictive value. With the ONN, we obtained an AUC of 0.870, an accuracy of 85.3%, a sensitivity of 75.0%, and a specificity of 86.5%, which were compared favorably with those of the CNN. The detection results of the small ANNs in Table 1 were selected by comparing the test results among 45 fully-connected ANN models[12]. As shown in Table 1, the diagnostic performance of the ONN with a sigmoid activation function was comparable to that of the selected ANN models and was obviously better than that of the ONN with the RELU activation function.

Predicting the glottic location

Table 2 reports the glottic location prediction results in the laryngeal images in the test set for the selected ANN[9], CNNs, and ONNs, with a sigmoid activation function for all the nodes except the nodes in the convolution part of the CNNs. In Table 2, a predicted box overlapping with the glottis, which is marked in white, was considered an accurate prediction, and the adjacent prediction was defined as one of the eight adjacent boxes of the predicted box overlapping with the glottis, which is marked in white (see Figure 3b)[9]. With the ONN, the accurate prediction and the adjacent prediction rates were 70.5% and 20.5%, respectively, which was compared favorably with that of the CNN. The prediction results of the fully-connected small ANN in Table 2 were selected by comparing the training results among 84 ANN models[9]. As reported in Table 2, the prediction accuracy for the glottic location of the ONN was between those of the CNN and the selected ANN model.

Discussion

A new deep-learning structure for image localization, called an ONN, was presented and applied to detect the location of pneumothorax in chest X-rays, resulting in an AUC of 0.870, an accuracy of 85.3%, a sensitivity of 75.0%, and a specificity of 86.5%. We also applied ONNs and CNNs to predict the location of the glottis in laryngeal images and achieved accurate prediction and adjacent prediction rates of 70.5% and 20.5%, respectively, with the ONN. The ONN was compared favorably with that of the CNN, a commonly used deep-learning structure for image recognition, and was compared decently with that of the selected ANN model[9,12]. Compared with a CNN, an ONN required only approximately 10% of the computations using a CNN to train images with an input resolution of 256 × 256 pixels.

An ONN extracted well the spatial location information of the input images by setting the same weight factor to the connections of the input nodes in a given column to a vertical layer and setting the same weight factor to the connections of the input nodes in a given row to a horizontal layer. This approach would be similar to using the latitude and longitude values on a map and finding the intersection. Having different biases for the vertical and horizontal nodes can help greatly in extracting the spatial location information. However, significantly increasing the number of vertical and horizontal layers will not be very important to improve extracting the spatial location information.

CNNs have shown excellent performances in classification and object detection for images through abstraction extraction from the images while generating the featured maps using several filters. However,

extracting the spatial location information required for localization may be slightly different from the abstraction extraction. Therefore, an approach from a different perspective, such as ONN, is required.

In addition, the diagnostic performance of the ONN with a sigmoid activation function for all the nodes outperformed the ONN with RELU activation function for all the nodes other than the output nodes, as shown in Table 1. Since the back-propagation method[4–6] uses the gradient descent, the RELU function greatly simplifies the process of calculating the derivative of the activation function, as seen from Eq. (1). However, an essential characteristic of the activation function within an ANN node should be to generate a signal that is symmetric in both the forward and backward directions. The closest form to this function is the sigmoid function. Although the RELU function looks similar to a sigmoid function in terms of shape, it might remove some important data flow information from the ANN[12] [refer to Eq. (1)] and cannot produce a symmetrical signal, which is an essential characteristic of the activation function within an ANN node. Therefore, the performance of the ANN cannot be optimized with the RELU activation function.

Fully-connected small ANNs have achieved excellent results in image localization as in previous studies[9,12]. However, to change the input image resolution for the ANN and the number of hidden layers, several dozens of ANN models must be individually trained, and the test results of these models should be compared to find the best model. This approach requires considerable time, effort and computing resources, even for small ANN models. This study showed that an ONN can be used as a quick selection criterion to compare small ANN models for image localization, since the ONN performed well compared decently with the selected ANN model, and training an ONN on images with an input resolution of 256 × 256 pixels requires a similar amount of computing resources as a small ANN to train the same number of input images with a resolution of 30 × 30 pixels.

In conclusion, as a new deep-learning structure for image localization, an ONN, which is simple, efficient, and completely different from a CNN, was applied to detect the location of pneumothorax in chest X-rays and to predict the location of the glottis in laryngeal images. Since the ONN extracted the spatial location information of the input images better than the CNN, its localization performance was compared favorably with that of the CNN. The ONN with a sigmoid activation function for fully-connected hidden nodes outperformed the ONN with the RELU activation function, which does not produce a symmetrical signal, an essential characteristic of the activation function within an ANN node. An ONN can be used as a quick selection criterion to compare the test results of small ANN models for image localization to choose the best model from several dozens of ANN models. Finally, our approach can accurately predict locations in medical images, reduce the time delay in diagnosing urgent diseases, and increase the effectiveness of clinical practice and patient care.

Declarations

Acknowledgments

This work was supported by Hanyang University, Seoul, Republic of Korea (202000000002924).

Author contributions

Study concept and design: J.S.K.; data preparing: Y.C. and J.S.K.; computer software programming and deep-learning execution: J.S.K.; data analysis and interpretation: J.S.K., Y.C., and T.H.L; manuscript drafting: J.S.K. and Y.C.; and manuscript revision: J.S.K., Y.C., and T.H.L.

Competing interests

The authors declare that they have no competing interests.

Data availability

The NIH Chest X-rays dataset used in this study can be downloaded from https://nihcc.app.box.com/v/ChestXray-NIHCC. The laryngeal images used in this study are not publicly available due to the confidential nature of the clinical imaging data and the potential risk of personal information leakage, but can be made available by the corresponding authors on reasonable request.

References

1. Gulshan, V. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402-2410 (2016).
2. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115-118 (2017).
3. Ehteshami Bejnord B. *et al.* Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* **318**, 2199-2210 (2017).
4. Werbos, P. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. *Ph. D. dissertation, Harvard University* (1974).
5. Sathyanarayana, S. A gentle introduction to backpropagation. *Numeric Insight* **7**, 1-15 (2014).
6. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536. https://doi.org/10.1038/323533a0 (1986).
7. LeCun, Y., Bengio, Y. & Hinton, G. E. Deep learning. *Nature* **521**, 436–444. https://doi.org/10.1038/nature14539 (2015).
8. He, K., Zhang, X., Ren, S,, & Sun, J. Deep residual learning for image recognition. *Proc IEEE Conf Comput Vis Pattern Recognit* Las Vegas, NE, USA, 770–778 (2016).
9. Kim, J. S., Cho, Y. & Lim, T. H. Prediction of the Location of the Glottis in Laryngeal Images by Using a Novel Deep-Learning Algorithm. *IEEE Access* **7**, 79545–79554. https://doi.org/10.1109/ACCESS.2019.2923002 (2019).
10. Lee, J. Y., Kim, J. S., Kim, T. Y. & Kim, Y. S. Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm. *Sci Rep* **10**, 20546.

https://doi.org/10.1038/s41598-020-77441-z (2020).

11. Kim, M., Kim, J. S., Lee, C. & Kang, B.-K. Detection of pneumoperitoneum in the abdominal radiograph images using artificial neural networks. *Eur J Radiol Open* **8**, 100316. https://doi.org/10.1016/j.ejro.2020.100316 (2021).
12. Cho, Y., Kim, J. S., Lim, T. H. *et al.* Detection of the location of pneumothorax in chest X-rays using small artificial neural networks and a simple training process. *Sci Rep* **11**, 13054. https://doi.org/10.1038/s41598-021-92523-2 (2021).

Tables

Table 1. Test results for detecting the location of pneumothorax in chest X-rays by the type of deep-learning methods, i.e., artificial neural networks (ANNs), convolution neural networks (CNNs), and orthogonal neural networks (ONNs).

[a]Cho *et al.*[12]; using a fully-connected small artificial neural network (ANN) with a sigmoid activation function for all the nodes with an input resolution of 20 × 20 or 30 × 30 pixels. [b]Cho *et al.*[12]; using a convolution neural network (CNN) with a sigmoid or RELU activation function for the fully-connected hidden nodes with an input resolution of 256 × 256 pixels. [c]This work; using an orthogonal neural network (ONN) with a sigmoid or RELU activation function for all the nodes other than the output nodes with an input resolution of 256 × 256 pixels.

Method	Hidden nodes	AUC	Cut-off	Sensitivity %	Specificity %	PPV %	NPV %	Accuracy %
ANN[a], 20 × 20	49	0.876	0.122	78.3	84.2	37.5	97.0	83.6
ANN[a], 30 × 30	49	0.881	0.101	81.0	83.6	37.4	97.3	83.3
ANN[a], 20 × 20	49-49-49	0.876	0.084	82.0	80.7	34.0	97.4	80.8
ANN[a], 30 × 30	49-49-49	0.882	0.101	80.6	83.0	36.5	97.2	82.7
CNN[b], Sigmoid	49	0.861	0.119	76.6	81.1	33.0	96.6	80.6
CNN[b], Sigmoid	49-49	0.859	0.128	76.9	82.6	34.8	96.7	81.9
CNN[b], RELU	49	0.829	0.072	80.8	76.9	29.7	97.1	77.3
CNN[b], RELU	49-49	0.795	0.134	73.9	82.6	34.0	96.3	81.7
ONN[c], Sigmoid	49	0.870	0.132	75.0	86.5	40.3	96.6	85.3
ONN[c], Sigmoid	49-49	0.866	0.091	80.4	80.5	33.3	97.1	80.5
ONN[c], RELU	49	0.840	0.088	74.6	82.5	34.1	96.4	81.6
ONN[c], RELU	49-49	0.820	0.072	78.1	78.4	30.5	96.7	78.4

Table 2. Prediction results for the glottic location in laryngeal images by the type of deep-learning methods. %, (number of images)

[a]Kim *et al.*[9]; using an artificial neural network (ANN) with an input resolution of 30 × 30 pixels. [b]This work; using a convolution neural network (CNN) with an input resolution of 256 × 256 pixels. [c]This work; using an orthogonal neural network (ONN) with an input resolution of 256 × 256 pixels.

Method/number of hidden nodes	Accurate	Adjacent	Inaccurate
ANN[a]/98	74.5% (149)	21.5% (43)	4.0% (8)
CNN[b]/49	68.5% (137)	25.5% (51)	6.0% (12)
CNN[b]/49-49	66.0% (132)	28.5% (57)	5.5% (11)
ONN[c]/49	70.5% (141)	20.5% (41)	9.0% (18)
ONN[c]/49-49	67.0% (134)	22.5% (45)	10.5% (21)

Figures

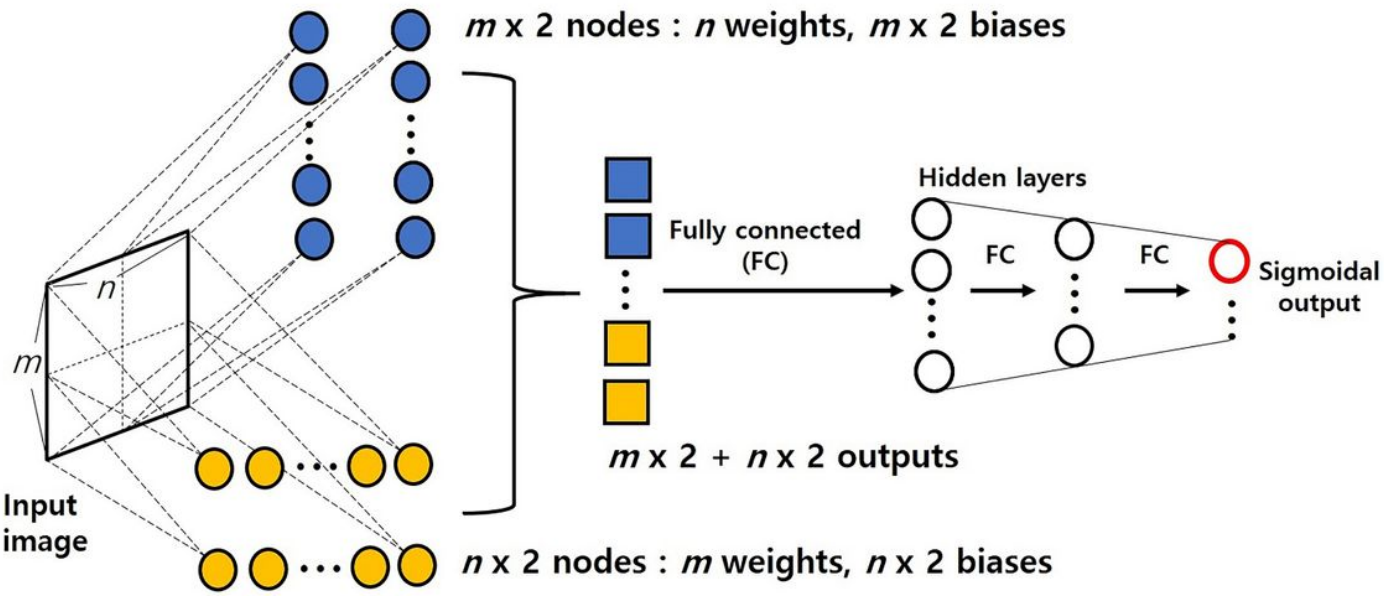

Figure 1

Schematic representation of the ONN structure.

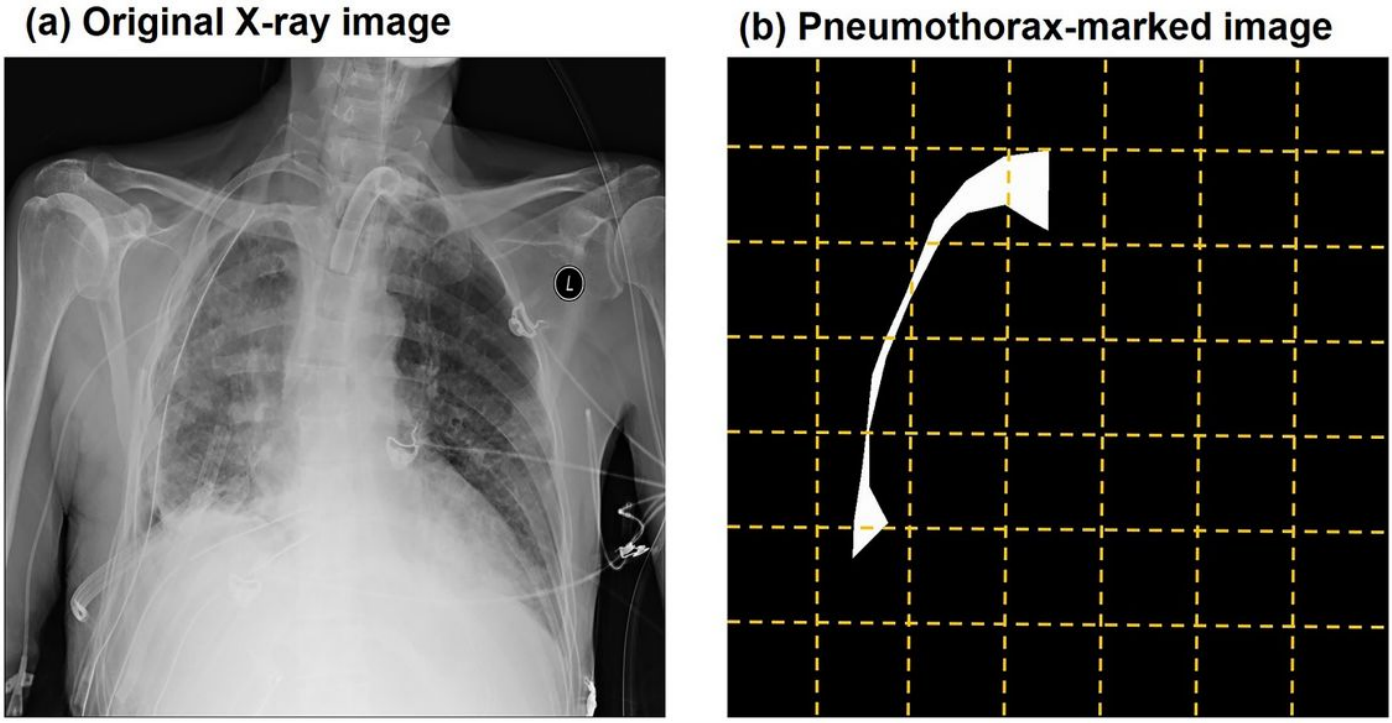

Figure 2

Example of chest X-ray image processing for application to an ONN: (a) Original X-ray image, (b) Pneumothorax-marked image.

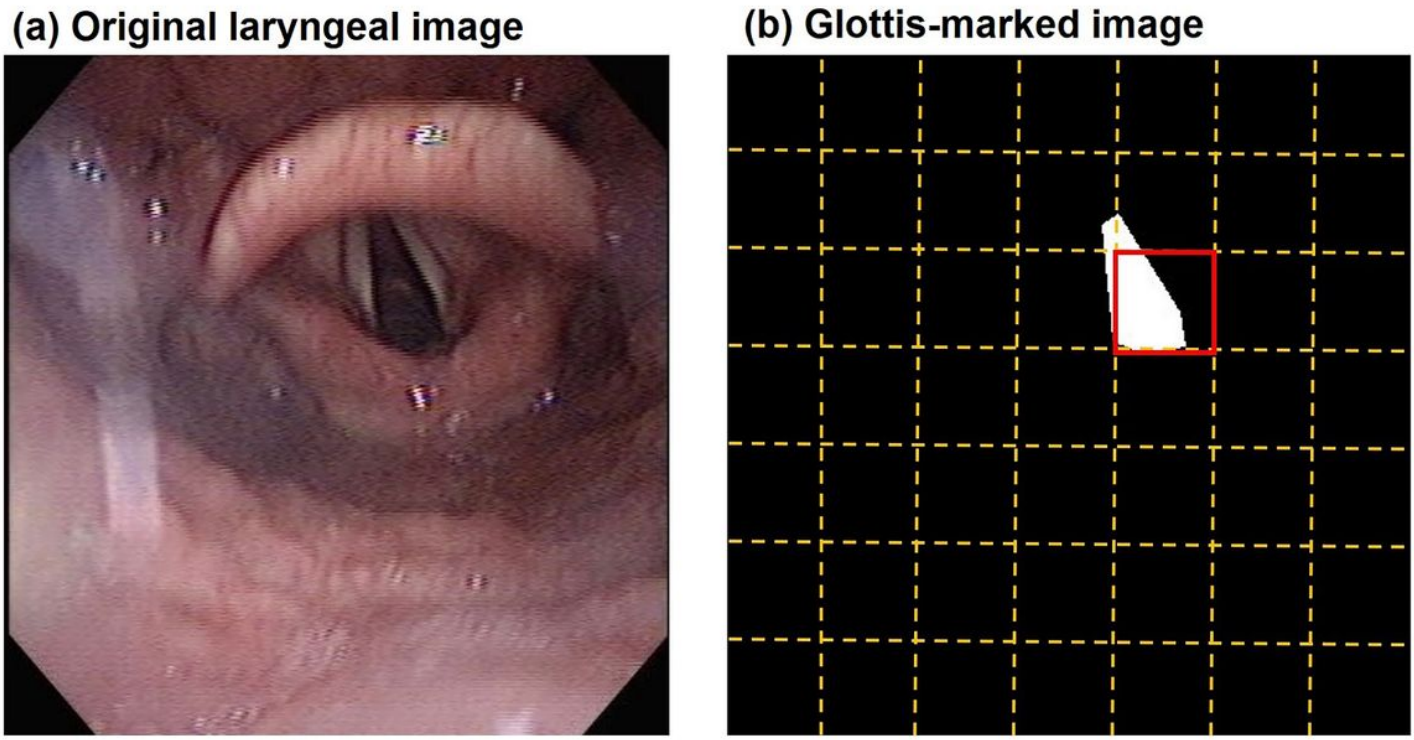

Figure 3

Example of laryngeal image processing for application to an ONN and CNN: (a) Original laryngeal image, (b) Glottis-marked image.

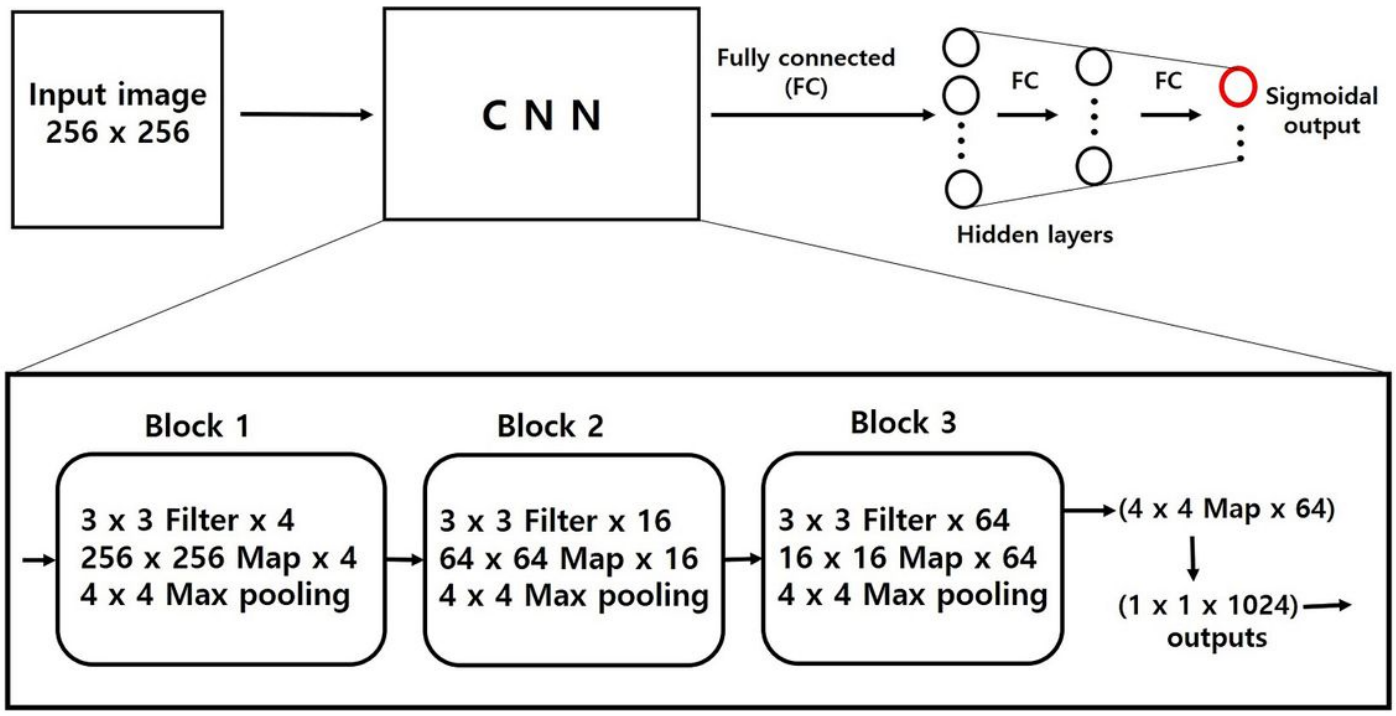

Figure 4

Schematic representation of the CNN structure.