Automatic Medical Images Segmentation Based on Deep Learning Networks

Ghada A. Shadeed¹, Mohammed A. Tawfeeq² and Sawsan M. Mahmoud³

¹,²,³Computer Engineering Dept., College of Engineering, Mustansiriyah University, Baghdad, Iraq

¹Corresponding author, e-mail: ghada.shadeed@gmail.com¹, drmatawfeeq@uomustansiriyah.edu.iq², sawsan.mahmoud@uomustansiriyah.edu.iq³

Abstract. In recent years, radiography systems have become more used in medical fields, where they are used for diagnosing many diseases. The size of the radiographs differs, as well as the size of the body parts for each patient. So many researchers crop the radiographs manually to facilitate the diagnosis and make it more reliable. Currently, the trend toward deep learning was commended where the deep learning proved its effectiveness in many fields, especially in the medical field, in which it achieves good results in diagnosing the most types of diseases. Deep learning performance increases significantly when the training process is focused on the region of interest. In this paper, segmentation is implemented by used deep learning model on the thoracic region of the radiograph in order to be cropped later. The proposed model provided automatic cropping of the radiographs where a semantic segmentation network is provided by Vgg19 model. A comparison is done with semantic segmentation network provided by Vgg16. The segmentation based on Vgg19 model outperforms Vgg16 model in cropping Chest x-ray images dataset automatically and quickly.

Keywords: Radiography systems, deep learning, semantic segmentation network, Vgg19, Vgg16

1. Introduction

Worldwide, around 450 million people per year suffer from pneumonia, it is a more common disease [1]. The radiation images of each patient are different in terms of the size of the body's organs, textures, and the structural shape of the persons [2]. In computer-aided diagnosis, this difference might make it difficult to diagnose diseases and lead to waste time. For diagnosing diseases such as chest diseases, the focus is on the area of the chest. In this case, there is no need for the presence of the rest of the body, such as the neck or abdomen in the radiological image, these details increase the training time and also reduce the accuracy of training. In addition, a lot of memory is needed during the process of training. So many researchers turned to crop radiographs manually before entering the diagnostic models. Image segmentation for medical images is the most common task currently. It has a great advantage in diagnosing and analyzing medical images. Segmentation is the initial stage within clinical applications [3]. In computer vision, TextonForest and Random Forest were common techniques in the semantic segmentation of images before use deep learning models [4, 5]. The segmentation technique is more accurate and efficient using deep learning models [6, 7].
In [8] an automatic crop method for the chest radiographs was proposed using adaptive binarization, which applied for pre-processing the images where Jaccard index is $0.941\pm0.032$. Also, standardizing is applied for pre-processing the images where the Jaccard index is $0.898\pm0.076$.

Recently, there has been a great success in deep learning in the recognition of handwritten numbers, figures, and objects in the picture as well as whole images classification [9] [10] [11] [12]. It differs from machine learning where machine learning algorithms are designed to learn by first understanding the data that has been labeled. Then more labeled data is used to get more output. However, when actual output is not required, they must be retrained through human intervention. Deep learning networks do not require any human intervention because the overlapping layers of neural networks place data through the hierarchies of different concepts, which ultimately learn through their errors [13].

Interest has increased about semantic pixel-wise labelling [14] [15] [16]. Some modern approaches tried to adopt deep architectures directly as these methods were designed to predict the class to pixel-wise labeling [14], where the results were encouraging, but it was coarse [17]. So semantic segmentation network (SegNet) [18] was used. SegNet has been designed to be high-efficiency architecture for pixel-wise semantic segmentation.

In this paper, encoder-decoder Convolutional Neural Network (CNN) was used which is one of the models of deep education called SegNet [18]. SegNet was used to set low-resolution features for input accuracy to classify pixel-wise to produce useful features for an accurately localizing boundary. The initialized weights of SegNet was from Vgg19 network [19]. Vgg is an abbreviation for Visual Geometry Group, which is a deep convolutional network for objects recognition and 19 refers to 19 layer of deep learning network. In this model, the convolutional layers are identical to the encoder network in SegNet. Meanwhile, the layers of fully connected were removed to make the encoder network of SegNet easier and take less training time than other recent architectures [20] [21] [22] [23]. Then, the fully connected layers are replaced with a decoder network. At the end of the decoder network, each layer in this network has corresponded to the encoder layer with pixel-wise classification. After obtaining a result from the SegNet segmentation process using Vgg19 model, the chest area is cropped later in many diagnostic processes. In addition, a comparison is made with SegNet using Vgg16 network [24] where the training time for the proposed model is lower, and the prediction accuracy is higher than the latter. In Figure 1, the architecture of SegNet is described where the RGB image is inserted into the encoder-decoder CNN to implement its segmentation at the end.

The remaining paper is organized as follows: Section 2 contains the proposed network and its structure. The proposed model’s results were presented and compared along with the data set in section 3. Section 4 introduced the conclusion and future work.

Figure 1. Architecture of encoder-decoder convolutional neural network prepared by Vgg19 model.
2. Proposed approach

In this section, the proposed approach is presented to crop the unnecessary parts of chest radiographic images to save time, effort and required memory.

A model that supports segmentation was suggested in the cropping process. The image was segmented by making hand-made labels of the X-rays. Then, depending on the segmentation result, the image is cropped.

2.1 Vgg model

VggNet usually refers to a deep convolutional network for object recognition developed and trained by Oxford's renowned Visual geometry group (Vgg), which achieved very good performance on the ImageNet dataset. Vgg network was introduced using convolutional layers stacked on top of each other at increasing depths. By max pooling, the volume size is reduced. Then followed by two fully connected layers, each with 4,096 nodes then followed by a softmax classifier. Vgg in all its layers uses very small convolution filters (3 × 3) and the convolutional step is equal to 1 pixel to reduce the number of parameters in this deep network [25].

There are two models of VggNet: Vgg16 and Vgg19.

A) Vgg16

Vgg16, involving 144 million parameters, contains 16 convolutional layers with very small receptive fields 3x3), five max-pooling layers of size 2x2 for carrying out spatial pooling, followed by three fully connected layers, with the final layer as the soft-max layer [29]. ReLU activation is applied to all hidden layers. The model also uses dropout regularization in the fully connected layers. Vgg16 trained on more than a million images from the ImageNet database. The network can classify images into 1000 object categories, such as keyboard, mouse, and pencil [30]. Vgg16 model takes an input image of size 224*224*3 (RGB image). The size of the Vgg exceeds 533 MB for Vgg16 due to the number of fully connected nodes and deep. Vgg16 Neural Network consists of 16 layers of deep neural network. Vgg16 network has less weight [26].

B) Vgg19

Vgg19 is a CNN that is trained on more than a million images from the ImageNet database [26]. The network can classify images into 1000 object categories, such as keyboard, mouse, and pencil. Vgg19 model takes an input image of size 224*224*3 (RGB image). The size of the Vgg exceeds 574 MB for Vgg19 due to the number of fully connected nodes and deep [27]. Vgg19 contains 19 layers of deep neural networks. Vgg19 network has more weight (138M weights and had 15.5M MACs [28]). A schematic of the Vgg16 and Vgg19 architecture trained on the ImageNet database is shown in Figure 2.

Figure 2: Architecture of Vgg16 and Vgg19 [31].
2.2 SegNet deep learning model

A SegNet (encoder-decoder CNN) deep learning model was used for segmentation and Vgg19 model pre-trained models were used to obtain the initial weights and layers to build SegNet model. The layers of Vgg19 are used to configure the five encoder layers and the decoding layers are identical to the encoder layers, as illustrated in Figure 1. After each convolutional layer, it is followed by the batch norm and ReLU. The training parameters were explained as follows: initial learning rate is 0.001, a minibatch size is 2, and Stochastic Gradient Descent with Momentum (SGDM) is used as the optimization algorithm for training.

Chest x-ray 14 dataset [32] and local data is used in this work to validate the proposed approach. The size of Chest x-ray images dataset is 1024 x 1024. The images then go through the pre-processing stage, where they are converted from a 2D image to a 3D image as required by the model. In the next step, the images are resized to 224 x 224, this step reduces the used memory which was one of the problems confronted during the work according to the specifications of the computer used. Following the pre-processing stage, the resized images are presented to the proposed model. The proposed model is based on the encoder-decoder CNN structure on the weights and layers of the pre-trained Vgg19 model retraining. The model classifies each pixel by using classification pixel layer where the pixel was split during the labeling process to two colors, the first is the black, which represents the chest part, and the white color, which represents the rest of the body parts other than chest part. The classification process was treated with two classes depending on the color of each pixel. This called pixel-wise labeling [14]. At the end of the training process, the image's segmentation is resulted, in which only the chest part will be in the black color.

Finally, the output is passed to the crop process to retain the chest portion of the image which represents the region of interest (RoI), while neglecting the other parts. The general structure of the work is illustrated in Figure 3.

3. Experimental result

In this section, the experiments were carried out using the database containing 3000 patients. 100 Patients were randomly taken to train and test the model. The proposed model was first applied to 25 chest-ray images and then to 100 images from the Chest x-ray 14 dataset and the local data (3.41MB image’s size and 410 KB label’s size), were used for the evaluation of the proposed model. Local data was taken from Sheikh Zayed Hospital, which is ten images, due to the difficulty of obtaining a permit.
and the lack of possibilities to extract the stored data. For training and testing, 60 and 40 images were used respectively with a special label made manually for each image as illustrated in the samples shown in Figure 4. The model is implemented in Matlab 2018b using neural network toolbox model for VGG-16 and VGG-19 network [24] [19], working on a computer with 16GB memory and Intel(R) Core (TM) i7-8550U CPU @ 1.80GHz.

The proposed model is programmed using the MATLAB program according to the following steps: Layers are first invoked, then classes are defined, followed by the categorization step each pixel in the image. The required network layers are then created for semantic segmentation. Followed by removing the existing pixel classification layer and adding a new pixel classification layer and connecting it to make it able to handle new classifications. This is followed by a network training step based on network classes and training data. Finally, the network is tested.

Jaccard index was adopted in order to compare the proposed model with radiological technologists. Jaccard index range rings between [0-1], and the higher the value, the better the results are, which means the two areas are more similar. The definition of the Jaccard index between two regions M and N is determined as follows [33]:

$$J = \frac{M \cap N}{M \cup N} \quad \ldots (1)$$

A Jaccard index represents the ratio of Intersection area to the union area, or Intersection-over-Union (IoU). In other words, a Jaccard index is used to measure the overlap ratio of each class.

The accuracy of the test dataset using the Jaccard index was computed as shown in Table 1, while the Jaccard index for each class was computed include IoU and the BF (Boundary F1) contour matching score between the predicted segmentation in prediction and the true segmentation in that we manually made as depicted in Table 2.

**Table 1.** Global accuracy of the test dataset using the Jaccard index

| Global Accuracy | Mean Accuracy | Mean IoU |
|-----------------|---------------|----------|
| 0.97            | 0.965         | 0.92     |

**Table 2.** Accuracy of the test dataset using the Jaccard index for each class

| Class   | Accuracy | IoU   | Mean BFScore |
|---------|----------|-------|---------------|
| lung    | 0.95149  | 0.90807 | 0.60799       |
| remain  | 0.98195  | 0.93915 | 0.8336        |

Figure 5 shows the original image, manual label, and a sample of chest ray image after applying the proposed model. It shows an output of the working mechanism of the Jaccard index where the difference between the proposed label and the actual label can clearly recognize. As can be seen from this image, there is a similarity between the result of the proposed model and the actual ones. The Jaccard index scores for the chest area is 0.95405 and for the remaining area is 0.94605.
The output of the proposed model (Vgg19) was compared with SegNet that its weights and layers were prepared from pre-trained Vgg16 and its architecture is shown in Figure 6. In this paper, for the purposes of comparison, several tests were conducted. The data set containing 25 images (15 images as a training set and 10 images as the testing set). The results of the proposed model have a training accuracy of 88.80 % (where the number of iterations was 700 and the training time was 17 min. and 43 sec.) while in the Vgg16 model it was 86.90 % (iterations are 700 and the training time is 11 minutes and 8 seconds). Increased training time for the proposed model compared to the comparison model was due to the increase in the number of layers in the proposed model. From this experiment, it can be concluded that even with limited availability of dataset, the proposed model is able to achieve high accuracy. In the testing phase, the results show good performance as illustrated in Table 4. On another test, the dataset of 100 images (60 images as a training set and 40 images as a testing set) were taken. The accuracy of the training in the proposed model was 93.35% (iterations is 3000 and the training time is 1 hour and 5 minutes and 7 seconds), while in the vgg16 model the accuracy of the training phase was 91.11% (iterations is 3000 and the training time is 49 minutes and 22 seconds). Whereas the results showed a
significant increase in accuracy of both the proposed model and the Vgg model when increasing the number of data and the results were similar for both as shown in Table 5.

**Table 4.** Results of test dataset using the Jaccard index where 10 images were used

| Models          | Global Accuracy | Mean Accuracy | Mean IoU  |
|-----------------|-----------------|---------------|-----------|
| Proposed model: Vgg19 | 0.83671         | 0.82833       | 0.70035   |
| Vgg16 model     | 0.8321          | 0.82659       | 0.69494   |

**Table 5.** Results of test dataset using the Jaccard index where 40 images were used

| Models          | Global Accuracy | Mean Accuracy | Mean IoU  |
|-----------------|-----------------|---------------|-----------|
| Proposed model: Vgg19 | 0.97            | 0.965         | 0.92      |
| Vgg16 model     | 0.97            | 0.965         | 0.92      |

![Figure 6. Architecture of encoder-decoder CNN prepared by Vgg16 network.](image)

The proposed model is compared with the previous model [34], which used SegNet prepared by Vgg16, which used 154 images. The proposed model showed superiority over the previous model where the amount of similarity of the resulting segmentation compared to the original segmentation entered for the previous model was 0.9595 while the proposed model achieved a similarity amount of 0.965.

4. **Conclusions and future works**

In this paper, it was suggested that the chest area can be automatically cropped, where the cropping process is done based on the resulting segmentation of a semantic segmentation proposal using deep learning. An encoder-decoder convolutional neural network based on the vgg19 network is proposed for image segmentation. The obtained results showed high accuracy and promising performance. The proposed model was then compared with SegNet, prepared by Vgg16. The proposed model achieved higher performance and accuracy compared with the vgg16 model when limited dataset are available. While similar results obtained when 100 images were taken as a dataset. This means that the proposed model works very efficiently when using a limited volume of training data and this, of course, helps to solve the problem of lack of large data. As future work, images that have been cropped through the model can be entered into a diagnostic model to identify the necessary diagnosis of the patient's condition.

**ACKNOWLEDGMENT**

We commend the efforts made to make the Chest X-ray 14 dataset available. We also thank the Emergency Radiology Department at Sheikh Zayed Hospital for providing local data.
References

[1] Ruuskanen O, Lahti E, Jennings LC, Murdoch DR. Viral pneumonia. Lancet. 2011; 377(9773):1264-75. http://dx.doi.org/10.1016/S0140-6736(10)61459-6. PMid:21435708.

[2] Zhu H, Sun W, Wu M, Guan G, Guan Y. Pre-processing of X-ray medical image based on improved temporal recursive self-adaptive filter. In: Proceedings of the 2008 The 9th International Conference for Young Computer Scientists; 2008 Nov 18-21; Hunan, China. USA: IEEE; 2008. p. 758-63. http://dx.doi.org/10.1109/ICYCS.2008.360.

[3] I. Despotovic, B. Goossens, and W. Philips, “MRI segmentation of the human brain: challenges, methods, and applications,” Comput Math Methods Med, vol. 2015, p. 450341, 2015.

[4] J. Shotton, M. Johnson, and R. Cipolla, “Semantic texton forests for image categorization and segmentation,” in 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.

[5] J. Shotton et al., "Real-time human pose recognition in parts from single depth images," Commun. ACM, vol. 56, no. 1, pp. 116-124, 2013.

[6] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in Advances in neural information processing systems, 2012, pp. 2843-2851.

[7] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 4, pp. 640-651, 2017.

[8] Masataka Imura, Yoshito Tabata, Rikuta Ishigaki, Yoshihiro Kuroda, Yuki Uraniishi, and Osamu Oshiro, Automatic Cropping Method of Chest Radiographs Based on Adaptive Binarization, 35th Annual International Conference of the IEEE EMBS,Osaka, Japan, 3 - 7 July, 2013.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in CVPR, pp. 1–9, 2015.

[10] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR, vol. abs/1409.1556, 2014.

[11] Kaymak, S., Almezghwi, K., & Shelag, A. A. S. Classification of Diseases on Chest X-Rays Using Deep Learning. Advances in Intelligent Systems and Computing, 516– 523. doi:10.1007/978-3-030-04164-9_69, 2018

[12] Sullivan, D. P., Winsnes, C. F., Akesson, L., Hjelmare, M., Wiking, M., Schutten, R., Lundberg, E. (2018). Deep learning is combined with massive-scale citizen science to improve large-scale image classification. Nature Biotechnology. doi:10.1038/nbt.4225

[13] https://hackernoon.com/deep-learning-vs-machine-learning-a-simple-explanation-47405b3eef08

[14] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” IEEE PAMI, vol. 35, no. 8, pp. 1915–1929, 2013.

[15] N. Hft, H. Schulz, and S. Behnke, “Fast semantic segmentation of rgb-d scenes with gpu-accelerated deep neural networks,” in KI 2014: Advances in Artificial Intelligence (C. Lutz and M. Thielischer, eds.), vol. 8736 of Lecture Notes in Computer Science, pp. 80–85, Springer International Publishing, 2014.

[16] R. Socher, C. C. Lin, C. Manning, and A. Y. Ng, “Parsing natural scenes and natural language with recursive neural networks,” in ICML, pp. 129–136, 2011.

[17] C. Liang-Chieh, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” in ICLR, 2015

[18] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 12, pp. 2481-2495, 2017.

[19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556v6, 2015.

[20] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in CVPR, pp. 3431–3440, 2015.

[21] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in ICCV, pp. 1520–1528, 2015.

[22] W. Liu, A. Rabinovich, and A. C. Berg, “ParseNet: Looking wider to see better,” CoRR, vol. abs/1506.04579, 2015.
[23] S. Hong, H. Noh, and B. Han, “Decoupled deep neural network for semisupervised semantic segmentation,” in NIPS, pp. 1495–1503, 2015.

[24] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[25] ZHANG, Feng, and Bao-jiang ZHONG. "Image Retrieval Based on Fused CNN Features." DEStech Transactions on Computer Science and Engineering aics (2016).

[26] ImageNet. http://www.image-net.org

[27] https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-exception-keras/

[28] Md Zahangir Alom, Tarek M. Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Brian C Van Essen, Abdul A S. Awwal, and ijayan K. Asari, The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches, Alom, Md Zahangir, et al. "The history began from alexnet: A comprehensive survey on deep learning approaches." arXiv preprint arXiv:1803.01164 (2018).

[29] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, in: ICLR 2015. Presented at the 3rd International Conference on Learning Representations, San Diego, CA, 2014

[30] https://www.quora.com/What-is-the-difference-between-VGG16-and-VGG19-neural-network

[31] https://medium.com/coinmonks/paper-review-of-vggnet-1st-runner-up-of-ilsvrc-2014-image-classification-d02355543a11

[32] Wang, Xiaosong, Peng, Yifan, Lu, Le, Lu, Zhiyong, Bagheri, Mohammadhadi, and Summers, Ronald M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. arXiv preprint arXiv:1705.02315, 2017.

[33] Masataka Imura et al, “Automatic Cropping Method of Chest Radiographs Based on Adaptive Binarization,” 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013

[34] Lamin Saidy and Chien-Cheng Lee, “Chest X-Ray Image Segmentation Using Encoder-Decoder Convolutional Network,” IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2018