Evaluation of Transfer Learning with CNN to classify the Jaw Tumors

Ahmed Khalid Ismael
Northern Technical University\Mosul.
Eng.ahm2003@gmail.com

AbdulSattar Mohammed Khidhir
Assistant Professor
Northern Technical University
abdulsattarmk@ntu.edu.iq

Abstract
Artificial Intelligence” (AI) This term refers to the idea that the machines can perform human tasks. Recently, researchers, professionals and companies around the world introduce deep learning and image processing systems that can analyze hundreds of X-Ray and Computer Tomography (CT) images rapidly to speed up the diagnosis of medical image and help to contain them. Dental diseases analysis is among the most innovative research fields, offering diagnostic and decision-making facilities for a variety of diseases, such as oral and maxillofacial diseases. Inside this paper, we present a comparison of recent architectures of the Deep Convolutional Neural Network (DCNN) for the automatic classification of two diseases depending on transfer learning with finetuned using a pre-trained network (VGG16, VGG19). The proposed work was tested using a small scale X-Ray panoramic dataset containing 116 images (58 ameloblastoma and 58 Complex Odontoma). As a result, we can assume that the pre-trained network (VGG19) demonstrates highly satisfactory results with a rate of increase in the accuracy of training and validation. Unlike CNN, pre-trained network (VGG16) demonstrates less performance when a small image dataset is available.

Keywords: Convolutional Neural Network (CNN), Deep Learning (DL), Transfer Learning (TL)

I. Introduction
There are different types of medical images like panoramic X-Ray, CT (Computed Tomography), MRI (Magnetic Resonance Imaging), etc. which are useful for disease diagnosis. Deep learning algorithms have a vast number of uses for the medical image processing applications [1][2]. They are effectively used in troubles such as brain tumor segmentation, liver tumor segmentation, anatomic brain segmentation, kidney segmentation[3][4], and glaucoma detection[5]. CNN performs remarkably well in medical image classification problems (mass or natural classification of breast tissue, lung pattern classification, classification of lung nodule[6]. Analysis of medical images using deep learning algorithms has attracted considerable attention from many researchers. In dentistry, various types of x-ray are used[7]. The most widely utilized tools for diagnosing dental diseases are Orthopantomogram (OPG) and Radiovisiography (RVG) x-ray images. OPG image captures the upper as well as lower teeth. In contrast, RVG x-ray images are used to diagnose an individual tooth. To reach its conclusion, such artifacts in the x-rays require domain knowledge and experience. That process takes time and is tired. Machines can
do such tasks with ease, speed and precision. The aim of this study is to use transfer learning with fine-tuned for
CNN pre-training (VGG16, VGG19)[8] to perform dental disease classification for two networks and to compare
the result. A set of (OPG) x-ray images is input to the CNN. To train the CNN a library of labelled images (OPG) is
used. Label images are those images where the type of disease has already been classified. Analyzing the x-ray
images will determine the disease. In particular, we considered for the classification task two diseases known in oral
and maxillofacial, ameloblastoma and Complex Odontoma. Both diseases in the oral and maxillofacial areas are
classified as benign, locally aggressive tumors. The origin of that epithelium is odontogenic. Ameloblastoma and
Complex Odontoma signs and symptoms are typically painless and are generally found in oral panoramic
radiography. Classic clinical finding Asymptomatic jawbone swelling is [9]. Some Deep Learning technologies
include detection and classification of breast cancer, detection of lung cancer, detection of skin cancer etc. Although
these methods have shown enormous achievement in terms of medical imaging success, they require a big quantities
of data, which is not yet available in this field of applications. Because no medical imaging dataset is available, our
work will be using Transfer Learning with fine-tuned [10] for (VGG16, VGG19). This paper explains comparing
accuracy for two networks of pre-training based on a classification of two dental diseases.

II. Background

The Artificial Neural Network (ANN) is also known as the Neural Network (NN). The functioning of neural
networks is the same neurons in the human nervous system. ANN's architecture layered, as defined in Fig. 1. Each
layer has several nodes which have an activation function. Nodes are fully connected with the next layer. There are
weights correlated with such relations. Each neuron performs input and dot product of weight, and feed the output
forward to next layer. The architecture of ANN assumes layers which are fully connected (FC). When used with the
input being an image, these fully connected (FC) layers raise the number of parameters drastically. The massive rise in parameters is not only a memory problem, it can also cause overfitting. Overfitting refers to the
network's status when it learns the features on the training dataset completely but does not generalize on the test
dataset reasonably well. Hence ANNs are not performing well with the images as data input. The dataset for the
train is split into training and validation set. On the training set, the network is trained and evaluated on the
validation set then. The network tends to reduce error on the train. The error in the test first decreases and then
begins to increase after overfitting. The point where the minimum test error is selected to be the optimal network
setting.

Fig.1. Artificial neural network, with 4 layers in total. First one is an input layer, two other layers are hidden and the last one is an output layer.

Deep learning is highly efficient for problems relating to image classification. CNN's do not require extraction of the
features in advance. Designed model CNN can learn the features on its own. Convolution neural network is one type
of ANN that supposing images are inputs. In ANN, the number of parameters in the network is increased by fully
connected (FC) layers, but the CNNs contain several convolution layers, FC, Dropout etc. As convolutional layers
share the weights across the spatial dimension, it reduces the parameters and can thus easily be scaled to image
inputs where there are a huge number of features input. Deep CNN algorithms can automatically learn the
representations of hierarchical features and capture regional patterns from OPG images in their many convolutional
and hidden layers. Wang [11] reported that deep CNN algorithms with only two convolution layers and one fully connected hidden layer could effectively perform edge detection. Therefore, because deep CNN architecture has a functionally powerful advantage in solving the problem of detection, it was chosen for use in this study [11][12].

Transfer learning is consist of two-stage:

- Feature Extractor: As an extractor function, a pre-trained model trained on a large dataset is used. Such features are then supplied as custom network inputs. The pre-trained model learns the basic low-level features from the huge dataset.
- Fine Tuning: The pre-trained model can learn high-level features, specific to the large dataset. Fine-tuning is performed to make the high-level functions compatible with the training data. Weights are unfrozen on certain layers and that these layers are used for training. This is done to make the pre-trained model more relevant to the training results. The weights of all the convolutional layer are frozen, but the weights of the last two fully connected layers are fine-tuned.

A CNN architecture model typically consists of five layers: “Input layer, convolution layer, pooling layer, full-connection layer and output layer”, as shown in Fig. 2. Also, it is known that the CNN model can be trained end-to-end to allow the extraction and selection of features, finally, classification.

The CNN architecture for our experiment has the following:

- Input layer: X-Ray images are the inputs in our experiment. The image dimension (200x200) is defined with the parameters.
- Convolutional layers: a convolution is a linear operation consisting of multiplying a set of input with weights. It is built for two-dimensional input; multiplication occurs between a two-dimensional weight array (filters) and input data array. We have 3 layers in the proposed architecture with a 3x3 scale filter and zero paddings.
- Pooling Layers: Pooling layers is a technique for accessing sample feature maps by summing up the existence of features in feature map patches. There are two types of methods of pooling, which are average pooling and max pooling. We've used Max pooling in the proposed architecture to calculate the maximum value for each feature map in each patch. For a stride of 2, the max-pooling is set to 2x2.
- ReLU layers: we have used ReLU layer for each convolutional layer
- Inner-product layers or fully connected layers: Treat the input data as a simple vector, and generate a single vector output. In this model, we have one layer of the inner-product. The last, a fully connected output layer with sigmoid activation.

Fig. 2. Main architecture of baseline CNN
In a CNN the convolutional layers derive the characteristics. The last two fully connected layers work as a neural network by classifying the functions and analyzing the images. Only the parameters in the last two fully connected layers of the CNN are changed during transfer learning (training with your images). In the convolution layers, the parameters are kept at the values obtained during the pre-training process. The aim is that the convolutional layers have learnt to extract the features. To recognize the features, we simply have to teach the last two layers that function as a neural network. Since features are geometric forms, it is not required to utilize the images from the eventual system with transfer learning for the pre-training. It is only essential for CNN to have learned all the geometric forms possible. This can be achieved by training on a comprehensive and diverse dataset.

III. Methodology

A. Dataset Collection and Augmentation

A large labelled dataset is essential to train the CNN. No labelled dataset is available for the dental diseases listed above. So, with the help of dentists and radiologists, the small dataset consisting of 116 images was labelled. Dataset cannot be made available for reasons of privacy and the scarcity of those pictures. To minimize overfitting as a consequence of learning from the limited datasets, this work utilized data augmentation to increase the size of the image samples by utilized the horizontal flipping and rotate by 20 before the data was fed to the machine learning. By horizontally flipping and rotate all of the images, the size of the training and test datasets was tripled. Due to practical constraints, the images used in the current study also are cropped and resized to 200,200 pixels. Fig. 3 shows augmentation.

B. Using VGG16 and VGG19

The reason for choosing the architecture of VGG16 and VGG19[8] was that it was widely accepted and regarded as state-of-the-art in both general and medical image classification tasks[1]; It has also been trained on large-scale data sets so that a transfer learning approach for large-scale image recognition can be adopted. A convolution neural network (CNN) architecture proposed by Zisserman and Simonyan VGG (Visual Geometry Group) in 2014 and utilized to win the ILSVR(ImageNet) competition in 2014 (K. Simonyan and A. Zisserman, 2014). Instead of having a large number of hyperparameters, the key feature of this architecture is that they focus on simple 3x3 kernels in convolutional layers and 2x2 sizes in max-pooling layers. Ultimately it's got two FC (Fully Connected Layers) followed by a softmax for output. The most common VGG models are VGG16 and VGG19, respectively containing 16 and 19 layers. The difference between VGG-16 and VGG-19 is that in each of the three convolutional blocks, VGG-19 has one layer more[13] as shows in fig.4. This architecture is perfect for deep learning and is very efficient in solving problems related to object detection and image recognition in complex non-medical image data[8], [12]. However, since it is important to classify black-and-white OPG images of different size, a possible problem with the VGG-16-19 network architecture was thought to be that learning efficiency could be reduced without correction, and the risk of overfitting could be significantly increased. Then, we adjusted the VGG-16-19 network architecture by changing the number of convolutional and hidden layers and hyperparameters, including the number of epochs,
batch size, loss function, optimizer, momentum and learning rate, to minimize overfitting as much as possible and to promote successful deep learning performance[14].

Fig. 4. Schematic diagrams of the VGG16 and VGG19 convolutional neural networks (CNN) architectures

C. Model Training
By using pre-processing, splitting and data augmentation techniques, our training dataset quantity is increased and ready to be passed along with the proposed models to the feature extraction stage to extract the correct and relevant features. Using an 80 to 20 percentage, divide the OPG radiographic image dataset into a training data set and validation dataset. The training and validation datasets were used directly for analyzing the OPG based on Keras framework in Python. Using transfer learning with fined tuned pre-trained network (VGG16, VGG19) to train the network. The weights for VGG16-19's Convolutions Layers have been frozen and the last two layers of FC (Fully Connected) used for training. The idea of transfer learning to be used as fine-tuning is to let the model learn from a large dataset some basic low-level features (e.g. edges and curves) of the image, so using the high-level features to be more specific to the training data. This is done to improve the model's adaptability to the training data. The two models are trained for 100 epochs with the batch size 32 and = 0.0001 learning rate. The final output layer used softmax as a classifier for the OPG.

IV. Results
This section discusses the findings of the oral and maxillofacial image classification. Before discussing these findings, let us describe the two most used parameters of the state of the art deep learning and computer vision: Train curve is calculated from the training dataset which indicates how well the model is learning, while the validation curve or test curve is calculated from a hold-out testing dataset which gives an indication of how well the model is generalizing. Additionally, the loss of training and validation is described as a summation of the errors in validation or training sets made for each example. The loss is not a percentage, as opposed to accuracy. To summarize, the best model is a model that generalizes well which is neither overfitting nor underfit.
The above figure presents accuracy and loss curve of VGG16. Indeed, from the epoch 1 to epoch 7, the accuracy curve of train data is quickly increasing where is equal to 9.5% then it decreasing to a value of 9.333%, and then in epoch 11 increasing to a value of 9.833 to become 10 in epoch 14 to the epoch 100. but the validation curve is quickly increasing from epoch 0 to 7 then become uncritical from epoch 8 to 25 and then stable with some drops to become 8.333 % to the last epoch. Observed decreasing of loss curve from epoch 0 to epoch 6 but the curve becomes unstable and starts to increase from epoch 24 to the last epoch. The pre-training model VGG16 classify 20 of 24 images for the testing image.

As it is shown in the figure above, the curve of train data (test data) can be divided into two intervals: the first one starts from epoch 0 to epoch 3 where we can observe a decrease of accuracy where the accuracy is equal to 5.833 % to 4.167 %. The second interval the accuracy starts to increase and converges to 8.333 % and become unstable until the epoch 24 it becomes stable at 8.750 % to the last epoch. For the loss curve of train and test data, we can observe a rapid decreasing. From epoch 0 to epoch 4 where is equal to 0.5115 then the train it starts to decrease until the epoch 100 where is equal to 1.7027e-04. But the test or validation curve start increasing to become 0.6269 and stay fixed from epoch 83 to 100. The pre-training model VGG19 classify 21 of 24 images for the testing image.

V. Conclusions
A cutting-edge neural network technology, called deep learning, has been applied in recent years to analyzing medical imaging and has shown a level of accuracy that is equivalent to or better than a clinician. This paper explains the effectiveness of transfer learning with fine-tuning in a deep CNN for the classification of oral and maxillofacial diseases in OPG images, in cases with limited training data and good results we used an OPG image dataset to apply two pre-trained VGG16 and VGG19 networks to discriminate oral and maxillofacial diseases. The experimental results indicated that the best overall accuracy of 87.50 percent was achieved by transfer learning with
pre-trained weights and fine tuning techniques for VGG19 compared to VGG16, which was 83.33 percent accurate. The findings presented indicate that the combination of appropriate deep CNN architectures and transfer learning techniques has effectively solved the issue of a limited image training dataset.

References
[1] G. Litjens et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, no. 1995, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.
[2] C. Park, C. C. Took, and J. K. Seong, “Machine learning in biomedical engineering,” Biomed. Eng. Lett., vol. 8, no. 1, pp. 1–3, 2018, doi: 10.1007/s13534-018-0058-3.
[3] W. Li, F. Jia, and Q. Hu, “Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks,” J. Comput. Commun., vol. 03, no. 11, pp. 146–151, 2015, doi: 10.4236/jcc.2015.311023.
[4] W. Thor, S. Kadoury, N. Piché, and C. J. Pal, “Convolutional networks for kidney segmentation in contrast-enhanced CT scans,” Comput. Methods Biomech. Biomed. Eng. Imaging Vis., vol. 6, no. 3, pp. 277–282, 2018, doi: 10.1080/21681163.2016.1148636.
[5] Y. Lu, J. Sun, and S. Ma, “Moving Object Detection Based on Deep Convolutional Neural Network,” Xitong Fangzhen Xuebao / J. Syst. Simul., vol. 31, no. 11, pp. 2275–2280, 2019, doi: 10.16182/j.issn1004731x.joss.1-FZ0368.
[6] B. Sahiner et al., “Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images,” IEEE Trans. Med. Imaging, vol. 15, no. 5, pp. 598–610, 1996, doi: 10.1109/42.538937.
[7] A. G. Farman and T. T. Farman, “A comparison of 18 different x-ray detectors currently used in dentistry,” Oral Surg., Oral Med. Oral Pathol. Oral Radiol. Endodontology, vol. 99, no. 4, pp. 485–489, 2005, doi: 10.1016/j.tripleo.2004.04.002.
[8] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
[9] S. Apajalahti, J. Kelpp, R. Kontio, and J. Hagström, “Imaging characteristics of ameloblastomas and diagnostic value of computed tomography and magnetic resonance imaging in a series of 26 patients,” Oral Surg. Oral Med. Oral Pathol. Oral Radiol., vol. 120, no. 2, pp. e118–e130, 2015, doi: 10.1016/j.tripleo.2015.05.002.
[10] A. Bhary et al., “Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images,” Pattern Recognit. Lett., vol. 129, pp. 271–278, 2020, doi: 10.1016/j.patrec.2019.11.013.
[11] Z. Chen, Z. Zhan, W. Shi, W. Chen, and J. Zhang, “Edge Detection using CNN,” vol. 1, pp. 603–612, 2016, doi: 10.1007/978-3-319-40663-3.
[12] W. Ouyang and X. Wang, “Joint deep learning for pedestrian detection,” Proc. IEEE Int. Conf. Comput. Vis., pp. 2056–2063, 2013, doi: 10.1109/ICCV.2013.257.
[13] Q. Zhang, H. Wang, S. W. Yoon, D. Won, and K. Sihari, “Lung nodule diagnosis on 3D computed tomography images using deep convolutional neural networks,” Procedia Manuf., vol. 39, pp. 363–370, 2019, doi: 10.1016/j.promfg.2020.01.375.
[14] N. S. Keskar, J. Noedal, P. T. P. Tang, D. Mudigere, and M. Smelyanskiy, “On large-batch training for deep learning: Generalization gap and sharp minima,” 5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc., pp. 1–16, 2019.