Convolutional Neural Networks for Medical Image Transfer Learning+ VLAD aggregation

Houssem FARHAT
Aix-Marseille Unversity
houssem-farhat@live.fr

Ronan Sicre
Ecole Centrale Marseille / LIF - QARMA
ronan.sicre@lis-lab.fr

Abstract

when we have an image classification problem, we directly think of a problem that needs to be solved by deep learning techniques using convolutional neural networks (CNN). In practice, Training a deep convolutional neural network (CNN) from scratch (with random initialization) is difficult, because it is relatively rare to have a dataset of sufficient size (Nima Tajbakhsh and Jianming Liang, 2016). That's why the researchers decided to use different techniques to solve this problem. In general, these techniques consist of using pre-trained networks on a large dataset, such as (ImageNet). Their idea was to transfer the knowledge acquired by this trained network and use it for new images for another classification problem. So, to apply transfer learning, there are two different strategies (ConvNet as fixed feature extractor, Fine-tuning the ConvNet), while in this paper we will use the first strategy and after we will try to introduce the Vector of Locally Aggregated Descriptors hoping to get à better results.

1 Introduction

Deep learning applications in medical image analysis can be traced back to the 1990s (W. Zhang and Schmidt, 1994)and(H.-P. Chan and Helvie, 1995), when they were used for microcalcification assisted detection in digital mammography. but at 1998 the world recognized the first ever convolutional neural networks, and what propelled the field of Deep Learning. This pioneering work by Yann LeCun was named LeNet5 (Y. LeCun and Haffner, 1998) after many previous successful iterations since the year 1988. In the years from 1998 to 2010 neural network were in incubation. At the time, there was no GPU to help with training, and even the processors were slow. Therefore, being able to save parameters and calculations was a key benefit. Most people did not notice their increasing power, while many other researchers progressed slowly. More and more data was available because of the rise of technology (cellphone cameras) to collect more data. And in the other side we have a new power of computing, CPUs were becoming faster, and GPUs became a general-purpose computing tool until we saw the revolution with Alexnet (Alex Krizhevsky and Hinton, 2012) which was a deeper and much wider version of the LeNet and won by a large margin the difficult ImageNet competition in 2012 thanks to the calculation made by the graphic card GTX 580 dedicated to the graphical calculation. and since 2012, every year we see another CNN coming out with new techniques to mark a new record on the competition imagenet. until we get to have CNN's that have a number of parameters that exceeds 100 million. And here when CNN have once again become a popular learning machine for various applications ranging from natural language processing to hyperspectral image processing and to medical image analysis. except that the problem arises when we want to train a convolutional neural network (CNN) from scratch, and this will necessarily take a lot of time and computing power. For that we will use transfer learning which consists of using a pre-trained model without needing to train it another time. However, this is not always the case. In fact, according to the type of images that we want to classify, we can choose which transfer technique we should use. For example if we have images that look like the images used for the training of the model, we can use fixed feature extractor then in the opposite case, will be better to use the technique of fine-tuning to adjust the model's weights to adapt with the new images. in parallel there are also other approaches based on Image vector representation to improve the classification of images, and these techniques had a lot of success, with the results obtained for the Identity documents classification using BOW,VLAD and Fisher vectors (Ronan Sicre and Furon, 2017) and (Jégou et al., 2010) In this paper we are interested to classify medical images while almost all the basic model is trained on images that belong to other categories. On the other hand our research consists at first in using the strategy fixed feature extractor to know what result we can obtain and to admit it like a base line and secondly how we can improve it by using the Vector of Locally Aggregated Descriptors (VLAD).

2 convolutional neural network (CNN)

Convolutional neural networks are currently the most efficient models for classifying images. Designated by the acronym CNN, they have two distinct parts. In input, an image is provided in the form of a matrix of pixels. It has 2 dimensions for a greyscale image. The color is represented by a third dimension, of depth 3 to represent the fundamental colors [Red, Green, Blue]. The first part of a CNN is the actual convolutive part. It functions as a feature extractor of images. An image is passed through a succession of filters, or convolution kernels, creating new images called convolution maps. Some intermediate filters reduce the resolution of the image by a local maximum operation. In the end, the convolution maps are laid flat and concatenated into a characteristic vector, called CNN codes. This CNN codes at the output of the convolutive portion is then connected to the input of a second portion, consisting of fully connected layers (multilayer perceptron). The role of this part is to combine the characteristics of the sub-sampling layer to classify the image. The output is a last layer with one neuron per category. The numerical values obtained are generally normalized between 0 and 1, of sum 1, to produce a probability distribution on the categories.

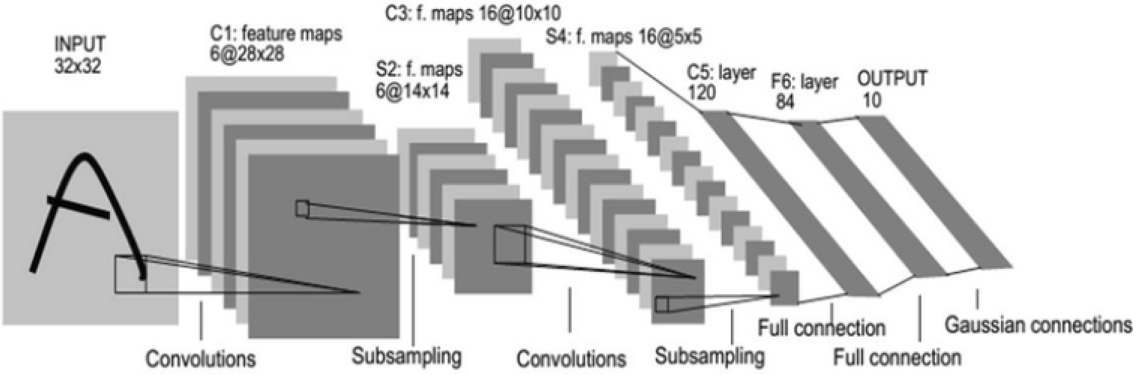


Figure 1: Full CNN Architecture

3 transferring representation

The fields of application of Transfer Learning are numerous. Mainly, knowledge transfer methods are very often used for image recognition as well as automatic language processing. These two areas of learning are very complex and time-consuming. This is why Transfer Learning brings a new breath to try to optimize these treatments by exploiting as much as possible models already trained.

3.1 Fixed feature extractor

A first method of learning transfer This model consists of using the neural network as a fixed extraction variable and applying it to our dataset.

steps to do this:

-We remove the last layer for classification from the network. (We can even remove the last 3 or 4 layers. In the end it is a choice to make as needed.

-We freeze the weights of the model and we use them as a fixed extract variable.

-The CNN codes are then extracted thanks to the fixed variable for all the images.

-A linear classification is performed for the new dataset with a Linear SVM for example.

4 VGG19

To perform classification, the networks used are all based on the VGG-19 model. VGG-19 is a convolutional neural network that is pre-trained on more than a million images from the ImageNet database. The network is composed of 19 layers, organized into 5 blocks of convolutional layers, and 3 fully connected layers,as shown in figure(2)

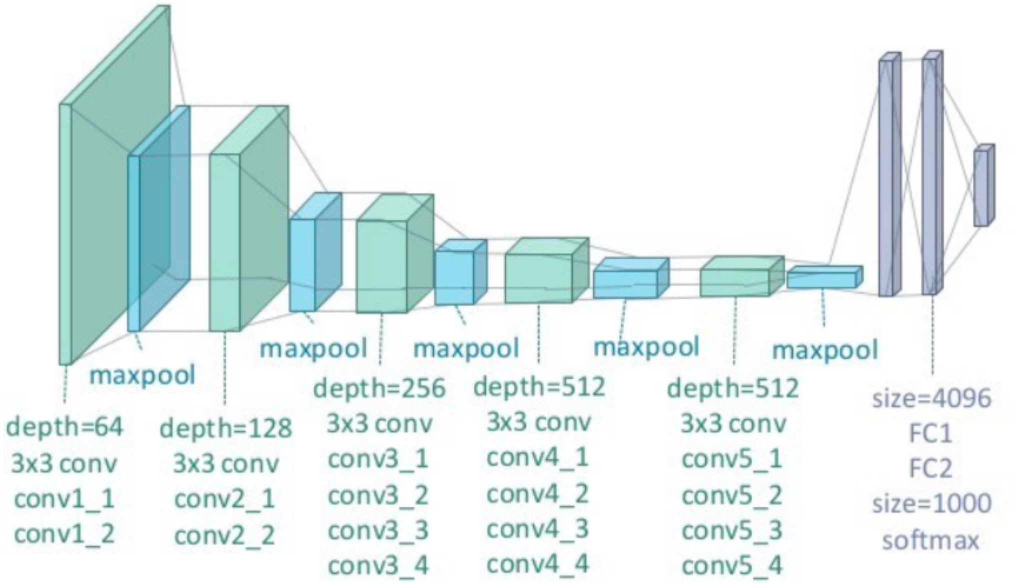


Figure 2: Full VGG19 Architecture

5 SVM

Support Vector Machine, or SVM, are supervised learning algorithms, which means that it is used to find a prediction function with annotated examples(KARI, 2018).

imagine a plane (two-dimensional space) in which two groups of points are distributed. These points are associated with a group: the points (+) for y> x and the points (-) for y <x. We can find an obvious linear separator in this example, the line of equation y = x. The problem is said to be linearly separable.

For more complicated problems, there is usually no linear separator. For example, imagine a plane in which the (-) points are grouped within a circle, with (+) points all around: no linear separator can correctly separate groups: the problem is not linearly separable. There is no separating Hyperplane.like shown in figure(3)

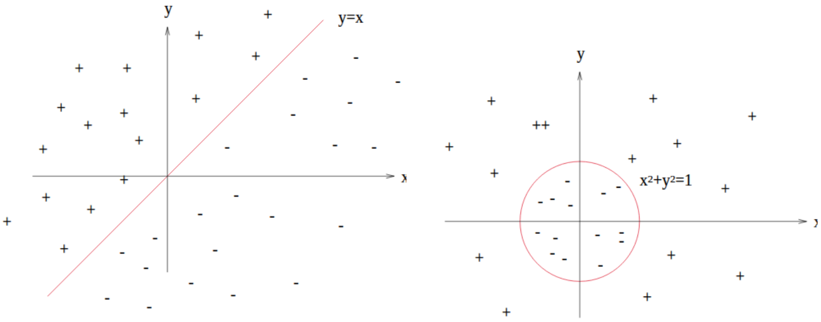


Figure 3: a linearly separable problem (left one) vs a not linearly separable problem (right one)

The margin is the distance between the hyperplane and the closest samples. These are the support vectors. Now we consider the case where the problem is linearly separable. Even in this simple case, the choice of the separating hyperplane is not obvious. There are indeed an infinity of separating hyperplanes, whose learning performance is identical (the empirical risk is the same), but whose performances in generalization can be very different. To solve this problem, it has been shown(Vapnik and Kotz, 1982) that there exists a unique optimal hyperplane, defined as the hyperplane that maximizes the margin between the samples and the separating hyperplane.

There are theoretical reasons for this choice. Vapnik(Vapnik and Kotz, 1982) has shown that the capacity of separator hyperplane classes decreases as their margin increases.

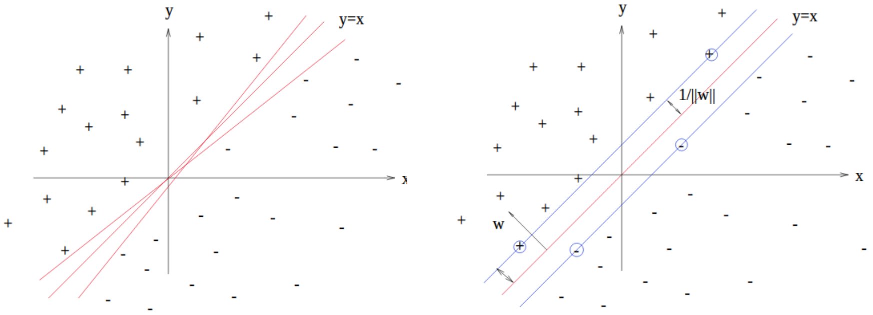


Figure 4: infinity of separating hyperplanes and optimal hyperplane maximizes the margin between the samples and the separating hyperplane.

6 Experiments

6.1 DATA

For the experiment we used 3 different datasets mini-MIT, xray-chest and kavasir.

6.1.1 Mini MIT data set

The Mini MIT, taken from here data set is a reduced data set from the MIT 67 scenes data set. It consists in two sets : a train set and a test set, and each of them contains three categories of images : bookstore,library and inside a bus. There are 40 images in each category for both sets.

6.1.2 Chest X-ray data set

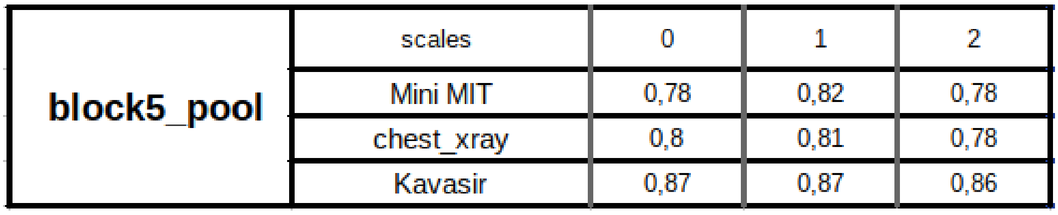
The Chest X-ray data set, coming from here, presents chest x-ray images selected from pediatric patients of one to five years old from Guangzhou Women and Children Medical Center, at Guangzhou. The data set contains 5 863 chest x-ray images which can be classified into two categories : those of pneumonia patients and those of healthy patients. we used 5 232 images for the train and 624 for the test.

6.1.3 Kvasir data set

The Kvasir data set, taken from here, consists in 8 000 images of the human digestive system. This data set is split into 8 categories : three categories of healthy patients : normal Z-line, normal pylorus and normal cecum, three categories of disease images : esophagitis, polyps and ulcerative colitis, and two categories related to the removal of polyp : dyed and lifted polyps and dyed resection margins. So we used 6400 for the train and 1600 for the test.

6.2 Results

As explained in subsection 4.1, all VGG19 network layers were extracted except fully connected layers. So we stop at the layer "block5_pool" which will be the output layer of the model. the experiment consist to use this model to predict the CNN code with different image sizes (224 * 224), ((224 + 320) * (224 + 320)) and ((224 + 320 * 2) * (224 + 320 * 2)) in input of the model. After we apply a average pooling with a normalization l2 to the new dataset and use a Linear SVM for the classification.



block5_pool	scales	0	1	2
	Mini MIT	0,78	0,82	0,78
	chest_xray	0,8	0,81	0,78
	Kavasir	0,87	0,87	0,86

Figure 5: Result as a baseline for different data sets and with a different input scale

Looking at the results we can see that the change of the input size can influence the performance of the prediction. On the other hand increasing of the input size does not bring back necessarily to have a better results.

6.3 VLAD: vector of locally aggregated descriptors

Now we start talking about how we can get a vector representation of an image that aggregates descriptors according to a locality criterion in a feature space.(Jégou et al., 2010) At first we can start with Learn a vector quantifier (k-means): c_1,...c_i,...c_k with c_i centroid of dimension d.

And after, for a given image :

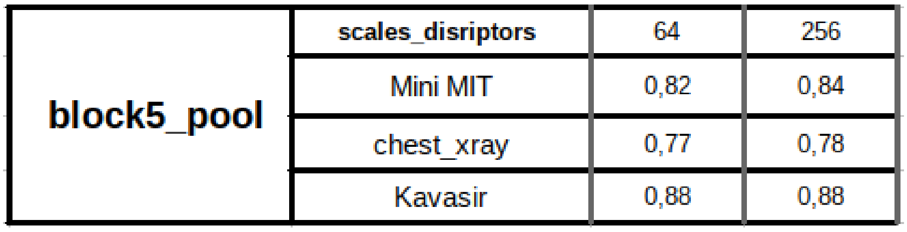
- assign each descriptor to closest center c_i
- accumulate (sum) descriptors per cell

v_i := v_i + (x_j- c_i)

measure repartition of vectors within a cell.

So we finish with a dimension D=k*d of our representation, with k typically between 16 and 256. In the end we shouldn't forget to apply L2-normalization to v vector and apply PCA reduction on VLAD can performed the results (Jégou et al., 2010).

6.4 Results



block5_pool	scales_disriptors	64	256
	Mini MIT	0,82	0,84
	chest_xray	0,77	0,78
	Kavasir	0,88	0,88

Figure 6: results of prediction after using VLAD+SVM with k=64/256

according to the results, we notice a remarkable improvement for the Mini MIT dataset. a little enhancement for Kavasir. and a degradation for Chest-xray, which shows that the application of VLAD may or may not improve the result. It all depends on a dataset.

7 Conclusions and prospects

This article addresses the problem of classifying medical images. We show that the CNN feature extracted from pre-trained networks can be successfully transferred to allow efficient and fast prediction without need to entrain the network from the scratch. We have also tried to apply a popular approach that produces a vector representation of an image from a set of local descriptors (VLAD). On the other hand we did not have a remarkable improvement for all the datasets. All this spur us thinking and trying to find the clogged element. For that we will have to try other layers like the layers fc1 and fc2 which are fully connected layers in the VGG19 network. Look for other datasets for the classification of medical images. Approach the problem using other Convolutional Neural Networks, for example the Resnet.

References

[Alex Krizhevsky and Hinton2012] Ilya Sutskever Alex Krizhevsky and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks.

[H.-P. Chan and Helvie1995] B. Sahiner K. L. Lam H.-P. Chan, S.-C. B. Lo and M. A. Helvie. 1995. Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network.

[Jégou et al.2010] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. 2010. Aggregating local descriptors into a compact image representation.

[KARI2018] Hichma KARI. 2018. Deep learning based medical image recognition.

[Nima Tajbakhsh and Jianming Liang2016] IEEE-Jae Y. Shin Suryakanth R. Gurudu R. Todd Hurst Christopher B. Kendall Michael B. Gotway Nima Tajbakhsh, Member and IEEE Jianming Liang, Senior Member. 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning.

[Ronan Sicre and Furon2017] Ahmad Montaser Awal Ronan Sicre and Teddy Furon. 2017. Identity documents classification as an image classification problem.

[Vapnik and Kotz1982] V. Vapnik and S. Kotz. 1982. Estimation of dependences based on empirical data, springer series in statistics.

[W. Zhang and Schmidt1994] M. L. Giger Y. Wu R. M. Nishikawa W. Zhang, K. Doi and R. A. Schmidt. 1994. Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network.

[Y. LeCun and Haffner1998] Y. Bengio Y. LeCun, L. Bottou and P. Haffner. 1998. Gradient-based learning applied to document recognition.

Annexe A. Titre de l’annexe