Review of Convolutional Neural Networks for Paedriatic Bone Age Estimation using Local and Cloud Environments

Blanca Hermosilla Campos[1], Marta Galván Prieto[1], Miriam Bautista Salinero[1], Sofía Vera Carretero[1]
[1] Biomedical Engineering Degree, Alcorcón Campus, Universidad Rey Juan Carlos.

Abstract

Bone Age Assesment is commonly used to control the optimal maturity of children bones as it can help the clinicians to asses possible genetic, endocrine or metabolic problems. Most of these examinations are carried out manually by the radiologists looking at an atlas reference. This implies a lot of time and a lost of precision due to each human judgment. To avoid this, in this paper a CNN architecture is proposed that will inform about the predicted bone age. The database that will be used to train and test the algorithm will consists on 12.800 images of the left hand and a csv document with the patient information. Due to the massive amount of data, several environments have been used achieving better results with a GPU in the cloud provided by Google Colab. It has been a limiter to the performance of the model. Better results could have arise if there were not cost barriers for the resources.

1 Introduction

Bone age assessment (BAA) is used in order to estimate maturity of a child's skeletal system since the difference between assigned bone and chronological ages may indicate growth disorders due to endocrine, genetic or metabolic problems. It is a common clinical practice in pediatric radiology for diagnostic and therapeutic investigations [1]. Alternatively, these examinations are used to optimize the time expend in interventions related to limb-length complications [2].

This technique is usually carried out by taking a X-ray image from the wrist to the fingerprints of the left hand, as it provides simplicity, minimum radiation exposure and the availability of multiple ossification centers[3]. Then, the most common process of examination is accomplished manually by a clinician through the Greulich and Pyle (GP) in 1959 or the Tanner–Whitehouse (TW) in 2002 methods. The GP method is the approach used by 76% of radiologists and is based on the comparison between the X-ray image and a reference atlas of representative ages. The TW system is based on a scoring system that examines 20 specific bones [2][4].

However, these techniques contain several disadvantages and challenges that should be considered. They require a large amount of time and the different subjective clinicians interpretations could affect the diagnose precision and the therapy decision[2].

Although software solutions, such as BoneXpert that uses a vision algorithm has been developed, they are still sensitive to the image quality[3].

As an alternative to this, deep learning models can be used as they have demonstrated optimal results in computer vision image tasks. Specifically, convolutional neural networks (CNN) have successfully worked in this type of applications such as segmentation and detection of patterns as they enable learning highly representative, layered, hierarchical abstractions from image data. In addition, many models attend to solve classification and regression problems such as bone age prediction.

Moreover, there are several studies that are already implementing this type of algorithms[2][3]. However, to the best of our knowledge, they did not use a big enough dataset of patients and images in order to achieve the best accuracy.

We propose a deep learning architecture made of CNNs that will asses doctors to perform a more accurate BAA that will lead to a better diagnose and a more personalized treatment. As our main goal is to train it as best as possible, a dataset with a very large number of images which varies in age and in gender is used.

On the other side, and as a consequence of the amount of data the study requires, the form of training the neural network will be considered. The purpose of this research also will include a qualitative and quantitative comparison between local and different cloud environments when making use of GPU in order to examine the one that performs the best with limited resources as a student. Hence, it will provide a practical reference and a source of information for future beginner researchers that will need to handle this massive amount of data in similar studies.

2 Methods

2.1 Dataset description and Exploratory Data Analysis

The RSNA Bone Age data set downloaded from Kaggle platform is composed of a training and a test sets containing 12,611 and 200 hand radiographs respectively. Images from the training set were labelled with the estimated skeletal age in months and sex by the experts at the time of the imaging. However, test images lacked of these labels.

For that reason, train test was used to generate two different subsets: training and test, each composed of 10,088 and 2,523 images respectively (Figure 1).

Out of the 10,088 images of the training set, 5,457 correspond to male patients and 4,631 to female's, ranging from newborns to teenagers. The mean age for males and females of the training set is of 135 and 118 months respectively. As for the test set, 1,376 images correspond to male patients with a mean age of 136 months whereas 1,147 images belong to female patients with a mean age of 116 months.

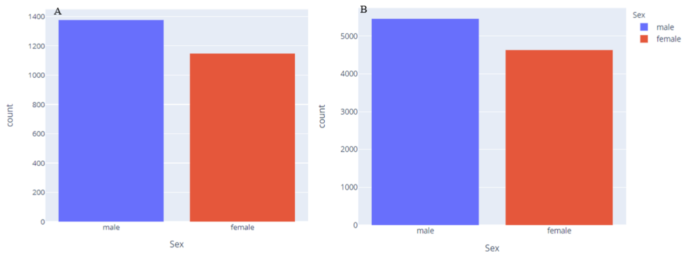

Figure 1: A. Male vs female patients in training set. B. Male vs female patients in test set.

Regarding the original train set, it includes 128 possible bone ages ranging from 1 to 228 months for both male and female patients (Figure 2).

Figure 2: Bone age distribution in months of the train set. Blue label corresponds to male patients and red label to female's.

Analysing the radiographs (Figure 3), it can be observed the variability of the data not only according to each skeletal age but also for the gray level intensities, image resolution and size.

2.2 VGG Neural Network

Tensorflow is a library provided by Google in order to be used for deep learning applications. From Tensorflow, Keras was used, which supplies a better framework to work with. From previous research done, VGG16 was found to be the best model available from Keras Tensorflow to use in this project.

VGG16 is a CNN model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" from ILSVRC 2014 competition.

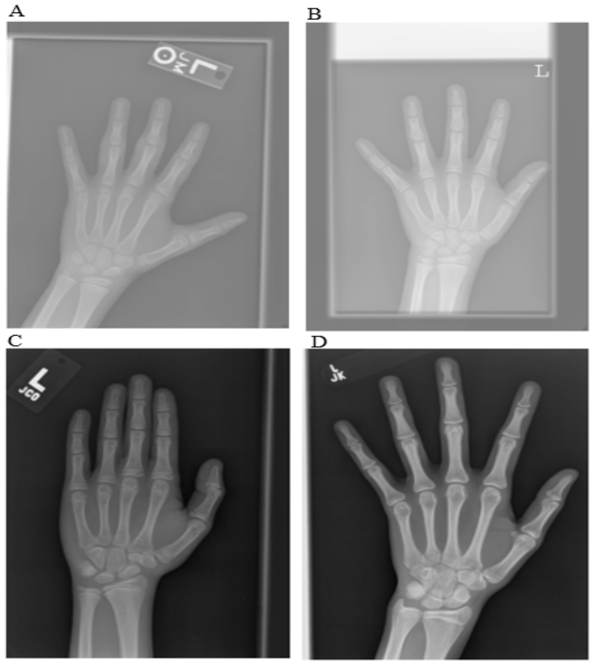

Figure 3: Comparison of different images according to bone age and sex. A. Male patient, 88 months. B. Female patient, 94 months. D. Male patient, 120 months. C. Female, 180 months.

The input to the convolutional layer 1 is of fixed size 224x244 RGB image. The image is passed through a stack of 16 convolutional layers, where the filters used are of size 3x3. Spatial pooling is carried out by five max-pooling, which follow only some of the convolutional layers. After every max-pooling the size is reduced to a half. The last two dense layers are a RELU activation function of 128, since there are 128 different age months as possibilities and a layer with a linear activation function which provides the actual prediction, respectively.

Furthermore, a batch normalization layer and early stopping were used in order to reduce the computational times and speed up the training time.

The input images to the model are preprocessed using a Data Generator which performs resizing and rescaling.

2.3 GPU for Deep Learning

GPU is the acronym for Graphical Processing Unit. They were originally developed for the purpose of speeding up video games, although later on, they were improved to execute other geometrical calculations. When the company Nvidia created a new design that implemented parallel computing, GPUs importance exponentially increased.

Since about 2008 every GPU provided by Nvidia is CUDA-enabled. CUDA (Compute Unified Device Architecture) is a general-purpose parallel computing and programming model that allows Nvidia GPUs to solve many complex computational problems in a more efficient way than on a CPU. CUDA-enabled GPUs dominate various applications areas, including deep learning. Deep learning researches rely on Nvidia CUDA Deep Neural Network library (cuDNN) which is a GPU-accelerated library that provides high implementations such as forward and backwards convolutions, pooling or activation layers. Among different widely used deep learning frameworks, Keras and TensorFlow use the cuDNN library for the acceleration of the neural networks computations [5].

2.4 Local and Cloud Environments

Local Environment

Jupyter Notebook is considered to be one of the most popular Integrated Development Environments (IDEs) for data science in Python [6].By default, Python Virtual Environment uses the computer's CPU to run the code. However, due to the heavy computing task involved in the training of the proposed deep model for BAA prediction, an environment with GPU was needed.

The Anaconda version that was already installed in the PC used for local computations corresponds to the 4.10.1 and the graphic card was the Nvidia GeForce GTX 1650. As a way to make Jupyter Notebook to run on the local GPU, the procedure has been as follows. Before all else, Visual Studio 2019 was installed, followed by the installation of the Nvidia CUDA Toolkit 10.1 and cuDNN SDK 7.6. Finally, a new Anaconda virtual environment was created with a Python version 3.8.5 and TensorFlow along other essential libraries and packages were installed.

Google Colab Cloud

Google Colab is a free cloud product developed by Google Research. It allows to write Python code into Jupyter notebooks in the web browser. It offers the possibility of sharing them with other people without the necessity of downloading, installing or executing any data in the local server[7].

Colab is very useful when dealing with artificial intelligence projects, data analysis and education. It allows free access to computational resources such as CPU, GPU and TPU. This is the reason why many times is the used resource to train models for Deep Learning. They usually take hours running into a CPU which can be the case of most of students that do not have a local machine with GPU. Hence, Google Colab could train the very same models in matter of minutes or seconds [7][8].

The types of GPUs available in Colab vary over time. The GPUs available from Colab often include NVIDIA's K80, T4, P4, and P100. There is no way to choose the type of GPU you can connect to in Colab. However, this resources by some means are limited. They cannot be always guarantee although the person is making use of the pro version, which implies a cost restriction [7].

The resources available in Colab vary over time to adapt to fluctuations in demand, general growth and other factors[7].

Kaggle Cloud

Kaggle is a subsidiary of Google LLC which started in 2010 allowing people to participate in data science and artificial intelligence competitions and that nowadays offers a platform of public data and a cloud-based workbench for data science and artificial intelligence education[9].

Kaggle provided free access to Nvidia K80 GPUs. Since 2019, its GPU has been upgraded to Nvidia Tesla P100, enabling a GPU to the kernel results in a high speedup during training of a deep learning model, as it happened in the previous cloud[9].

The exact speed-up varies based on a number of factors including model architecture, batch-size, input pipeline complexity, etc. That said, the GPU opens up much great possibilities in Kaggle kernels [9].

Anyway, as it happens with Google Colab, the resource is limited. They are implementing a limit on each user's GPU use of 30 hours/week [9].

Microsoft Azure Cloud

Microsoft Azure is a cloud service created by Microsoft to provide a mechanism in order to build, test, deploy and manage applications and services using data centers [10].

It includes multiple options of products in which we can find Azure Machine Learning Studio, the web portal for developers and data scientists for Azure Machine Learning [11].

For university students, it exists the possibility of making use of Microsoft Azure for Students which provides a credit of 100 dollars with limited benefits. For this reason, the only option that have been used is the Nvidia K80.

3 Experiments and Results

3.1 Model Results

In Figure 4 obtained from compiling the model, it can be seen the loss functions of train and validation sets, which are obtained by minimizing the mean square error. This metric was applied since it is the evaluation method most commonly used in regression models.

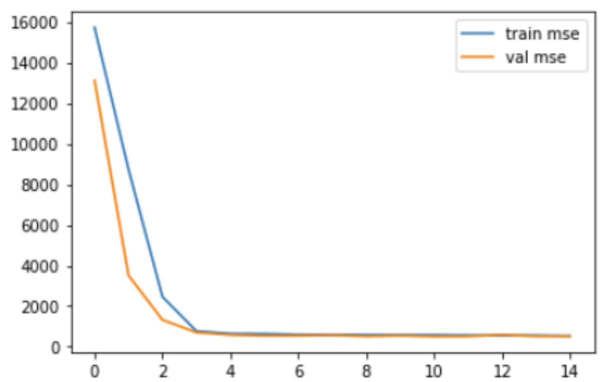

Figure 4: Mean square error validation and training graph.

It can be seen that both loss functions decrease in a similar way, but still obtaining high values. A well, the slopes decrease very fast in the first 5 epochs but remain constant for bigger values of epochs. Due to the high computational time required, we were not able to modify different parameters in order to improve the accuracy of the model.

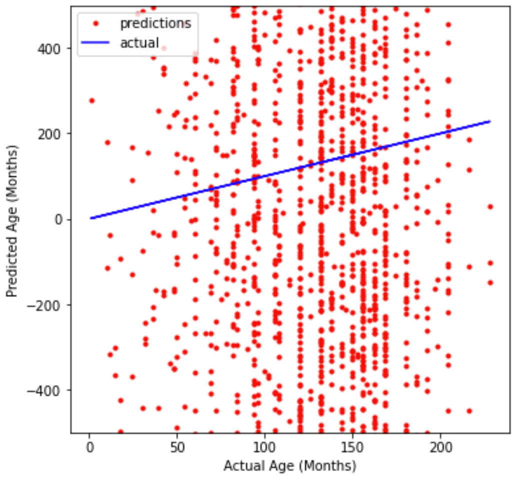

Figure 5: Mean square error validation and training graph.

Furthermore, when plotting the actual age against the predicted one, as it can be seen in Figure 5, it shows that the outcome is not as good as it could be.

3.2 Environment's comparison

Although the different previously mentioned environments were used for executing the neural network, problems still arose in each one of them. Independently of the high amount of data and that each environment uses a different processing unit, the common problem was the time needed to execute the model.

When talking about the data downloading, in the Local Server this was an extra issue. It required a high quantity of free space in the personal computer and increased downloading time. On the other hand, on Google Colab and Azure there was the need of using Kaggle's API for downloading and unzipping the images. Although this procedure was more convenient, it still required time. This step was omitted in Kaggle.

One of the first problems we encountered when using Azure was the different offers of nucleus when creating environments in different countries. For example, when the location was East France, only 6 nucleuses were available, against the 24 nucleus that were available when using Western Europe. This also implied a limitation of memory, which made it impossible to use an advanced GPU environment when using France's location. Furthermore, the only accessible GPU had a cost of 1.17$/h, which we had to keep in mind so that the maximum of 100$ was not reached. All the computations performed in azure made us consume 75% of budget. The economic affair was not an issue neither in the other Cloud environments nor in Local.

Even though Colab and Kaggle are completely free, they both presented time limitations. In the case of Google Colab, the maximum session time allowed was 12 hours, however, the experience led us to believe that the disconnection of the session happens before the time is reached. In the same way, Kaggle offered 35 hours per week, which was sufficient time for the computations. [12]

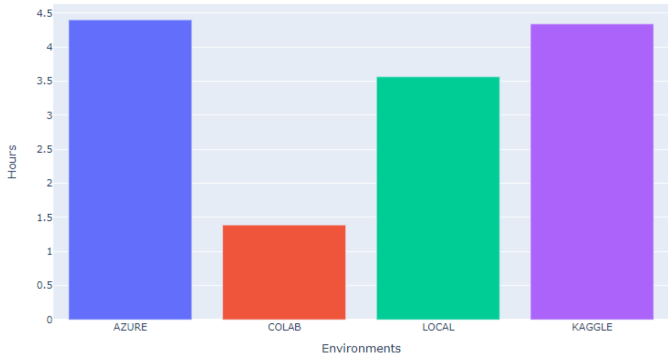

Figure 6: Comparison of computational times in different environments.

In Figure 6 a comparison of the computational time of all the environments is shown. For Azure, this was 4 hours and 40 minutes, approximately the same time that Kaggle took. Then, Local took 3 hours and 30 minutes and, finally, Google Colab only was 1 hour and 39 minutes, therefore being the fastest one.

4 Conclusions

Estimation of bone age in children is not only necessary for ensuring a correct development but it is also an indicator for the diagnosis of some diseases. Therefore, having a precise outcome is an important matter. Taking this into account, the proposed VGG16 model is not the most advisable one for this purpose. Nevertheless, we believe modifications could be performed in order to improve the results.

Each environment had advantages and disadvantages. Local environments are not usage restricted, however there are resource limitations from personal computers. At the same time, they are not as scalable as Cloud computing. Regarding the benefits Cloud environments offer against the Local, Cloud is proven to be more suitable.

Considering all the limitations we faced due to our student status, we are confident that Google Colab is the best option for this project. Its fast computational speed and its accessibility act in its favor. However, if we were to have a different Azure subscription, this environment would become a better fit. It will allow us to have a broader range of options when choosing what GPU to use, without having to worry about budget issues.

To sum up, all four alternatives are reasonable good for Big Data usage. Taking into account the characteristics of the chosen database within the available resources as students, we can conclude Google Colab presented fewer inconveniences. Furthermore, from all the difficulties we have encountered, we have reached the conclusion of the amount of resource limits there are in education, which links with the debate of how big companies are not sharing their capabilities at the speed at which areas such as deep learning and Big Data are increasing. In the same way, it is striking that education and research, two important fields in the development of new professionals, happen to be the most affected due to lack of funding.

Acknowledgments

Special thanks to Jose Luis Rojo and Sergio Muñoz Romero for teaching us the Massive Data Processing course.

5 Data availability

All the data used for this project is the one downloaded from Kaggle RSNA Bone Age Prediction page. The original dataset was published on CloudApp, but it was uploaded to Kaggle in order to better work with it. This data is publicly available and can be accessed using the following page: `https://www.kaggle.com/kmader/rsna-bone-age`

6 Code availability

The code is freely available at `https://github.com/miriambautistasalinero/MassiveDataProcessing`

References

[1] D. M. R. C.Spampinato, S.Palazzo, "Deep learning for automated skeletal bone age assessment in x-ray images.," *ScienceDirect*.

[2] L. H., "Fully automated deep learning system for bone age assessment. journal of digital imaging.," *Springer*.

[3] A. A. K. A. A. S. Vladimir I. Iglovikov, Alexander Rakhlin, "Paediatric bone age assessment using deep convolutional neural networks.," *Springer*.

[4] A. L. Creo, "Bone age: A handy tool for pediatric providers. american academy of pediatrics.," *American Academy of Pediatrics*.

[5] N. Developer, "Nvidia cudnn," 2021.

[6] P. H. Vasconcellos, "Top 5 python ides for data science," 2018.

[7] Google, "Preguntas frecuentes."

[8] Na8, "¿machine learning en la nube? google colaboratory con gpu!," 2019.

[9] Wikipedia, "Kaggle," 2021.

[10] Wikipedia, "Microsoft azure," 2020.

[11] o. Peter Lu, "¿qué es azure machine learning studio?," 2020.

[12] J. Hale, "Kaggle vs. colab faceoff — which free gpu provider is tops?," 2019.