# Application and usuability of neural networks in borders devices

Mateus Raimundo da Cruz
National Institute
of Telecomunications
Minas Gerais, Santa Rita do Sapucaí
37540-000
Email: mateusraimundo@mtel.inatel.br

Leandro Ribeiro
National Institute
of Telecomunications
Minas Gerais, Santa Rita do Sapucaí
37540-000

Felipe Agusto Pereira
de Figueiredo
National Institute
of Telecomunications
Minas Gerais, Santa Rita do Sapucaí
37540-000
Email: felipe.figueiredo@inatel.br

***Abstract*—The quantized neural network is a low precision and optimized neural network that tries to work in dispositive and low computer power devices that need an energy efficiency. This study test the QNN in a Raspberry Pi 3B in different scenarios a point some applications at this type of network. In add, this paper aims to point the good points and the problems in use in relation to the accuracy and the frame rate of this type of devices and network jointed.**

## I. Introduction

Did you notice when you finish watching some movie on some streaming service, in the next search of films that you want to see they recommend some movies in the genre that you like or similar that you just finish watching? Or when you searching for a new T-shirt to buy, the next few advertisements that you receive is that product or similar? This is artificial intelligence and machine learning working through the process of marketing. Artificial intelligence and machine learning is a study focused on how the computer can learning through the results of their predictions. Most of the areas today can use the machine learning process to improve some processes and help people to make better and more accurate decisions.

In addition, larger companies like Netflix, Google, Facebook, and others are be using these strategies for a lot of time to loyalty their users and make that more people come to his platform every day. But, a variety of areas can use machine learning, like Health, Advocacy, Engineer, and others. In health, for example, is possible to see algorithms that scan an X-Ray image and can inform to the professional some information and helps to separate what X-Ray that is urgent and there are no risk or low risk, this can improve the time of the professional and create a filter in all the exams that need to be ascertained.

Not only, but the machine learning techniques can also be combined with the concept of IoT (Internet of Things) and enlarge much more the applications. The IoT is a concept that looks to create a connection between some objects present in everyone's daily life and that creates the possibility to gather more data to transform into valuable information to his user. But, the large set of data creates a problem, what process all this data to make this valuable? And this is a work for machine learning to process and make predictions, classifications, regressions, and recognition.

But, another problem in merge IoT and machine learning is the divergence of the requisites of computing power. While IoT needs hardware of low cost to mass production, the machine learning technique normally uses a lot of computing power to make precise predictions, trains, and classifications. If combined an IoT dispositive with a traditional CNN, the speed of the process can be very slow or not useful in real applications, that need some accuracy and speed of the dispositive to make rapid decisions.

And for these cases, it was developed the QNN's (Quantized Neural Networks), that are neural networks with extremely low precision (e.g., 1 bit), weights, and activations. This is possible by the uses of neural networks in devices with low processing power like Raspberry Pi. But, the low processing causes a low precision in the recognition and classification of the target. And this paper aims to introduce the reader to some concepts like machine learning and artificial intelligence, IoT, QNN's propose a practical implementation of QNN's with Raspberry Pi and demonstrate, indicate some applications that the QNN can be implemented. In add, the final discuss the results and the benefits in uses of the Neural Network.

## II. Technologies

There is hardly an application in the real world with full use of machine learning, because IA is an umbrella of several concepts and technologies like computer vision, machine learning, robotics, Natural language processing, and much more. Because of this, this topic is exclusively for an explanation for all important concepts treat in this paper like computer vision, IoT, QNN's, and so on.

The joint work of these technologies and concepts is extremely important for applications of the real world. The project that will be discussed use computer vision with Quantized Neural Networks and concepts like IoT to create an application take can be used in several areas like the smart farm, security, utilities in home, and some areas of manufacturing.

### A. TensorFlow Lite

The main idea behind the development of TensorFlow Lite is the use of machine learning models in IoT and, mobile devices. The framework of the TensorFlow Lite is open source and possibles the use of deep learning in edge devices. In the site of TensorFlow is possible to see a variety of examples using the technique for the detection of objects, smart answers, and images classification with a large number of examples of implementation. [Site tensorflow]

### B. Quantized Neural Networks

Deep learning is capable of doing a series of different tasks such as object and speech recognition, statistical machine translation, Atari and Go games, and much more[1]. But all this requires a certain amount of computational power. For devices that do not have much memory or power alimentation, the Quantized Neural Networks (QNNs) is the perfect role.
The QNNs are neural networks with extremely low precision weights and activations. This reduces memory size and drastically the power consumption of the device [1]. This makes this machine learning technique perfect for devices like Raspberry Pi that are low memory, power and processor to become functional.
Some studies indicate that QNNs can reach 51% accuracy and seven times faster in an unoptimized CPU kernel. [1]

### C. Computer Vision

Is a field of IA that's try to identify things across images from cameras, the things that are considered targets can be trained across an algorithm to become recognized by the program. The computer vision can be used for a variety of purposes like securing, identifying peoples that are lost or be wanted by the law, recognized and offer access to peoples, and much more. For another example, computer vision can be used in agriculture to identify plagues, diseases and locate lost animals like cows or horses on the farm. But, the uses need to be checked for the use of QNN because the time to process and the frame rate per second (FPS) very slow in some devices.

## III. Applications

### A. Agriculture

Computer vision is already present in diverse applications like agriculture production, and industrial food production. And the process uses these techniques to classifying foods like orange, lemon, rice, and more and provides a subjective analysis of the process, making precise data and in this way is possible to automatize a greater number of tasks, providing more accuracy and more time to the workers realize other tasks. [1]
A lack of application in food production is about the use of computer vision to classifies the quality of the food or some errors or defect in the food, but this requires a complex system because the great numbers of a variable that system needs to work like: texture, color, format, size and more. Not only, but the actual necessity of a great environment to the system also imposes some restrictions of the applications, is good to the algorithm that the background of the subject needs to be standard, good quality of illumination in the subject (not less and not much), good equipment used by capture the information and etc. [2]
Not only, but some studies also discuss the adaptation of computer vision and machine learning techniques para validation of the quality of foods can be the major potential of observation of this technology, but this is necessary that the system can handle to automatic classify a great variety of product in great variability that can influence the results of the classification. [3]
However, one of the great premises of this area is the development of computer vision to use 3D techniques of visualization in colors to increase the accuracy of the system and the speed of the classification. [3]
Therefore, the project tested in this paper can be an important pass to reduce the cost of implementation of computer vision in small business, but the performance in classification of a quantized neural network counts and must be taken into account.

### B. Manufacturing

The technology of computer vision in common in manufacturing industries, but the union of machine learning techniques with computer vision can change the actual scenarios in the industries, and because of this, machine learning is one of the pillars in the fourth industrial revolution, the industry 4.0.
This revolution is a no back way, and machine learning can increase productivity and optimize the workflow process in the industry. Not only, but the fourth industrial revolution with artificial intelligence also can be used to make predictions about potential fails in the process and increase more efficiency in the industry. [4]
The IoT is another pillar in industry 4.0, and the main proposal of the project discussed in this document is to join concepts of IoT and Machine Learning together. Increasing the massive implementation of machine learning models to create more applications of technology.
But, is right to consider that machine learning technique is a hard task, but much problems (that will be discussed in this paper) like variations of weight, size, texture, format, color and the propriety nature of the three-dimensional world turns the subject a not easy target to classifies, because of this, the error tax is a problem. [5] Manly in neural networks adapted in borders devices like the used in this test.
But for small industries or industries that are beginning in the market, the price of the project and the applicability can be one of the most important aspects. Because the actual industrial applications computer vision can be not cheap and accessible for everyone.

### C. Surveillance and security

The actual system of surveillance is in the most part manual, but, the use of artificial intelligence to recognizes peoples in locals that are not permitted is a great use for a more simple model that only recognizes peoples and not identifying these peoples. For this, a quantized neural network can be used and a large number of devices can be spread in the area. This reduces the cost of implementation, and not only, possibly the actual security team to work in other tasks. In addition, it's evident that a lot of information present in the video can be processed by the actual technology, and is not necessarily higher resolutions like 4K to recognizes a person, the major challenges are in the analytical power of the actual power of computer vision and the conditions of light of the environment. [6]
But in surveillance at home, a small application with low cost and easy implementation can be an option, for moments that are nobody in the house or in periods of vacation. But not only, but these applications are also used in a store or on a farm is almost part of the time without anyone to guard.
Because of this, the test of this application is made by simulations of people recognize to attend the most part of this application, and provide some information about the results that the model gives.

## IV. A brief description of the project

The project has two main objectives, to integrate a deep learning model in edge devices, simulating the integration of machine learning techniques in IoT devices to operate together. And to test the performance in a variety of roles, like Smart Home, Manufacturing, Agriculture, Autonomous Vehicles, and so on. Like is explained in the next section, this project uses the microprocessor Raspberry Pi 3B+ and according to raspberrypi.org these model count with "1.4GHz 64-bit quad-core processor, dual-band wireless LAN, Bluetooth 4.2/BLE, faster Ethernet, and Power-over-Ethernet support (with separate PoE HAT)" and the entire configuration of the board can be viewed in fig. 1 above,[Site do raspberry pi], all the configuration offers by the Raspberry Pi 3B is sufficient to guarantee four to five frames per second in average to the model that will be used.
This becomes a limitation in some uses like Autonomous Vehicles and the test took demonstrates the risks and the accuracy and speed that are need in these cases to respond to the environment and to the things that comes to happen randomly like peoples entering in front of vehicles, balls that hist the sensor or camera and much more.

The entire specification of the device is shown below:

- Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz
- 1GB LPDDR2 SDRAM
- 2.4GHz and 5GHz IEEE 802.11.b/g/n/ac wireless LAN, Bluetooth 4.2, BLE
- Gigabit Ethernet over USB 2.0 (maximum throughput 300 Mbps)
- Extended 40-pin GPIO header
- Full-size HDMI
- 4 USB 2.0 ports
- CSI camera port for connecting a Raspberry Pi camera
- DSI display port for connecting a Raspberry Pi touch-screen display
- 4-pole stereo output and composite video port
- Micro SD port for loading your operating system and storing data
- 5V/2.5A DC power input
- Power-over-Ethernet (PoE) support (requires separate PoE HAT)

The project uses one Picamera too, that is a device capture and make recognizations in real-time by the processor and the algorithm. In all the applications discussed here is needed some cameras to take the information of the real world, process, and make decisions. The technical information about the Picamera is listed below:

- Shot: 1/4 5 M
- Opening: 2.9
- Focal distance: 3.29
- FOV: 65 degrees
- Type of sensor: Colour CMOS OminiVision
- Size of sensor: 3.67 x 2.74 mm (1/4 " format)
- Pixels: 2592x1944
- Video: 1080 p (30 fps)
- lens: f = 3.6mm, f/2.9

The algorithm allows using three resources like in source: webcam or camera, video, and photo. And is used a camera of cellphone Moto G4 to record the videos and to pass through in the algorithm, the purpose of using the cell camera is to create simulations where not is necessary to separate a source of energy to fee the raspberry, not necessary to start the dispositive in every local and don't need to move all the setup like keyboard, mouse or screen.
This impact of the use of another source like a cellular camera results in alteration of results because the camera gain FPS in video and turns the movement more smooth and are added in quality of the camera in the relation of resolution, sharpness, contrast, and colors.
The use of a webcam with good resolution with devices that improve the quality of the device is possible too with Coral USB for example. It's possible to use the Raspberry Pi 4 and gain ever more FPS on the real-time video, and not recording a video and passing through the neural network.
Are recorded a video using a Picamera too, but the video is made in a room with a low incidence of light, and because of this the video is not used in the tests of the experiments, is not possible to compare the two ways because the high differences on the two types of equipment, and I've decided to use only the camera to compare only differents locals, angles and levels of light to put in the test the neural network and

not the impact of the quality of equipment in the performance of the quantized neural network, that is extremely important but not the objective of the project.

## V. Setting-up the project

This section has to object to the description pass-to-pass of how the project is built, what devices and configurations have been made, and to transform all of this into a recipe.
This project use the microprocessor Raspberry Pi 3B+ but a large variety of devices can be used to replying this application, because the Tensor Flow Lite used in this project supports a various number of devices, including: Smartphones, devices with Linux Incorporated and other micro controllers. But have in mind, all the steps explicates in this article is for the setup described below:

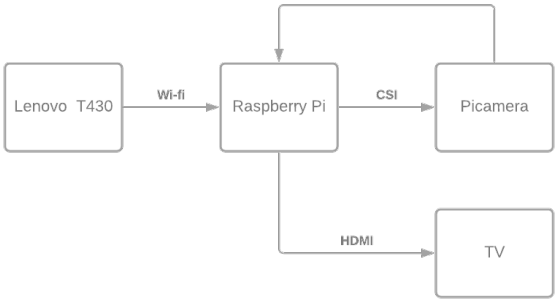


Fig. 1. Diagram of the connection between the devices

The complete setup consists: Raspberry Pi 3B+, Lenovo T430, Samsumg TV, internet onnection, Picamera (5MP) and some keyboard and mouse.
The first step is to set up all the physical connections correctly, make sure that all the necessaries components are available to use. To set up the Raspberry Pi, connect both the Keyboard and mouse in the USB ports, the HDMI between the Screen and the Raspberry Pi, Picamera in the camera serial interface (CSI) and the Power in the micro USB port of the micro processor and make sure that the power supply 3A and 5V for the device.
The second step in the project is to set up the Raspberry Pi operational system. The Noobs is the easy way to install the Operational System, here is used the Raspbian with GUI (Graphical user interface). To install the Raspbian in Raspberry Pi take a microSD card with minimum 8 Gigabytes of store capacity and insert in your machine (can be use a adaptor microSD x USB or a SD Reader with help of a adaptor of microSD to SD) and format the card. Go to the https://www.raspberrypi.org/ and download the pack of necessary files and extract in the formated cart, insert in the Raspberry Pi and install the operational system. For more instructions, consult the documentation in raspberrypi.org.
Now what the system has been installed, its time to make the connection between the notebook and the Raspberry across the internet. This is optional and can be solved with a good pair of mouse and keyboard, but in case you prefer to configure and control the Raspbian across your computer, follow these steps.
Was two ways to control the device through the internet. One is by the Virtual Network computing Software (VNC), and another is to Secure Shell (SSH), so the major difference is about the necessity of a Graphic User Interface (GUI) to take control of the Raspberry because SSH involves only prompt commands to control and navigate to the system.
The first step is to enable the VNC option in the Raspberry Pi control settings. To do this, go to the menu ¿ preferences ¿ Raspberry Pi configuration and in the interface tab select the enable option for the VNC connection, now is necessary that the VNC viewer software be installed in the machine, in the terminal ‘sudo apt-get update‘ to get any necessary update in the machine and after digit ‘sudo apt-get install realvnc-vnc-server realvnc-vnc-viewer‘ and wait at the process finish and well-done, the software is now in the Raspberry and can make the connection between her. To do this, open the software in the Raspbian machine and annotate the IP of the machine, in the computer, also open the VNC viewer and digit the IP. If everything is correct, will be asked for the user and the password, digit both correct and press enter, now it is possible to control the Raspbian System in Raspberry pi of the computer.
To user SSH in place of VNC, all you need to do is install PuTTy if the operational system is windows of use the terminal if the operational system is Linux. After all these steps, is now necessary to update the Raspberry Pi to make it possible to install the necessary packages that are needed to algorithm work. After all actualization, the download and installation of TensorFlow in the raspberry pi as needed and can be made by some commands with PIP.
The next step is to install the OpenCV to be possible to made detection of an object in images and this can be made with the library matplotlib too, but the OpenCV presents fewer errors and is easier to work. After installing the OpenCV and all the dependencies is necessary to compile and install Protobuf, that as a packet that implements Google’s Protocol Buffer data format and this can be by one line of code using apt-get install protobuf-compiler’ in superuser mode in Linux. Now, is necessary to create a directory and clone the repository to this destination. All the necessary files are encountered the GitHub and the suggested name for the folder is TensorFlow. After creating a directory and copy the repository for GitHub is time to create an environment for the program and install packet SSDLite-MobileNet.
Now, everything is completely installed and adjusted and after all the installation, the raspberry pi is now capable to do identification objects in webcam, video, and image, and the next steps are to made some tests with the project and visualize the performance of the algorithm.

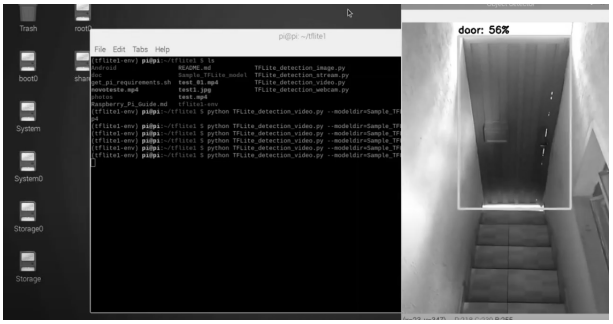


Fig. 2. Algorithm in use

## VI. Simulation

The simulation performance an important aspect of the project, because the utility and application of this project depend on what are the requirements of the application. Some tasks require less speed of the classification and some tasks need to be accurate and rapidly and the project cannot handle these requirements.

All the simulations are made with the camera of the smartphone Moto G4, and the reason for this is already explained in section IV. The simulation is made thinking in three different scenarios with different angles, the intensity of light, and the presence of objects. The more hard scenario thought to the algorithm are darker, environment with a higher level of detail and angle realist in the ceiling. And the easiest scenario for the algorithm is not much detail in the environment, the angle is realistic and at the same time, a clear view of the subject, and the light is perfectly uniform and constant, and the average scenario has a middle intense of the two and an angle of the ceiling too.

The maximum duration of the videos are thirteen seconds and are made three of this by simulating an installation of these project in a normal house to identifying peoples calling in front of peoples entering the house. The choice of angle and locals is a realist way used to install cameras of surveillance and to in the same time test, the performance of the camera in different scenarios with good, middle, and bad light, and in all the algorithm presents a good answer to the environment.

All three recorded videos after recorded are transfer to the raspberry pi and with the right commands, are passed in the algorithm to visualize t.he performance and to see all the errors that are made. The metrics to calculate the performance are discussed in the next section.

## VII. Results

To retrieve metrics about the performance of the algorithm is used some metrics of the evaluation of the classification made by the algorithm, and the metrics are:

- True Positive: When the algorithm classifies the subject correctly.
- True negative: When the algorithm classifies the subject as true but that shouldn't.
- False Positive: When the algorithm classifies the subject wrongly.
- False Negative: When the algorithm classifies the subject as false but that shouldn't

The error tax makes a relation of the false-positive plus false-negative classification divided by all classifications made and are used to demonstrate how a tax of errors the algorithm made.

*1) Error tax:*

$$Pe(y(x)) = \frac{FP + FN}{FN}$$

The precision corresponds to a proportion of examples of the class positive, correctly classifies em relation of all positive classes. That is, how classifications are made correctly in relation to all classifications correctly positive made correctly plus the classifications false-positive made and can be described by the formula below:

*2) Precision:*

$$Precision(y(x)) = \frac{TP}{TP + FP}$$

The recall is the tax of true positives, the recall corresponds to the proportion of examples of the positive class correct classified, and can be described by the formula below:

*3) Recall:*

$$Recall(y(x)) = \frac{TP}{TP + FN}$$

The specificity is given by the proportion of examples of the negative class that are correctly classified and can be represented by the formula below:

*4) Specificity:*

$$Specificity(y(x)) = \frac{TN}{TN + FP}$$

With all of these formulas and metrics in hands, the simulations are viewed and annotate all the classifications the algorithm has made, and with this, the taxes are calculated. The results are discussed simulation by simulation and the view of the author is expressed.

### A. Firt simulation


In this simulation, the presence of light is middle and has some objects in the video that can cause some confusion in the algorithm classifications. The video has thirteen seconds of time and is recorded around 11:00 AM, the sun is high and the shadow is hard, the local of recordation is a little dark and the person come off the light to the dark and this causes some explosion of bright. The metrics of this simulation is showed below:

*1) Error tax:*

$$Pe(y(x)) = 0,8973$$

*2) Precision:*

$$Precision(y(x)) = 0,9294$$

*3) Recall:*

$$Recall(y(x)) = 0,5278$$

*4) Specificity:*

$$Specificity(y(x)) = 0,55$$

The performance is good for a neural network with a general model for classification, the lack, and explosion of light is not a great problem to the algorithm and in general, the classification is correct, two majors errors made is to classify the person like a dog and some times the algorithm no recognize that is a person in the video and lose the classification, the others errors made in another object is a lack of a specialized model to recognizes people.

### B. Second simulation


This another simulation has made by trying a dark scenario with more details in the scenario and this can confuse more the classifications of the algorithm. But the results present a good performance of the algorithm, and this is a surprise because the darker scenario does not influence much in the classifications in this simulation. The result is shown below:

*1) Error tax:*

$$Pe(y(x)) = 0,3605$$

*2) Precision:*

$$Precision(y(x)) = 0,9761$$

*3) Recall:*

$$Recall(y(x)) = 0,5963$$

*4) Specificity:*

$$Specificity(y(x)) = 0,9090$$

In dark peoples with low detail of the information, the algorithm made a greater performance, classifying the object

correctly and in the major part of the time, this show what a quantized neural network can do in cheap devices.

### C. Third simulation


This simulation presents a great scenario for the algorithm, with a high incidence of light, great angle for the subject, and a low level of things and distraction in the environment for made classifications, and because of this, the algorithm can do great work in this work, even with a general model for classifications with eight subjects recorded in the model. The results below confirm this:

*1) Error tax:*

$$Pe(y(x)) = 0,1304$$

*2) Precision:*

$$Precision(y(x)) = 0,9859$$

*3) Recall:*

$$Recall(y(x)) = 0,8961$$

*4) Specificity:*

$$Specificity(y(x)) = 0,9687$$

The subject (person in the video) in recognizing correctly all the time of the video is recorded, and prove that a general model can do some jobs if the way of installation like angle to the subject and some aspects of the environment like light and little objects in the scenario can help a lot to the performance of the algorithm.

## VIII. Conclusion

The fault of a train to specialize the neural network made the results gone much worse than expected. Classifications of doors what coach and people what dogs are common in the test because of the lack of training and the use of a general sample offer by google.

The necessity of more FPS is a problem too because the model of Raspberry used in the training cause a mean of 4 to 5 FPS, and this is unacceptable for some applications that need a quick answer for the neural network. But, the final price of the application is less than $50,00, and the applicability of a computer vision system specialized to classified some subjet is a great relation cost versus benefit.

Concerning FPS, a newer model of microprocessor like Raspberry Pi 4B can easily handle some applications like surveillance tested in this video. But more tests need to be made to guarantee that.

More studies in relation to quantized neural networks need to be made, because exist a lack of information about these types of artificial neural networks, the more part of the research is appointed to the studies of Convolutional Neural Networks, Recurrent Neural Network and etc. But is hard to find studies that explain some uses in the quantized neural network, and this can be a problem for future studies. Therefore, the quantized neural network can play a great role in advanced IoT technologies, because the more number of bits is a positive aspect to the economy in the construction of devices with less processor power and economy of a battery of the IoT devices, making the prices of dispositive more low and the mass production acceptable.

## References


[1] N. V. G and H. K. S, "Article: Quality Inspection and Grading of Agricultural and Food Products by Computer Vision-A Review," *International Journal of Computer Applications*, vol. 2, no. 1, pp. 43–65, 2010. [Online]. Available: http://www.ijcaonline.org/archives/volume2/number1/612-863

[2] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Computers and Electronics in Agriculture*, vol. 153, no. August, pp. 69–81, 2018.

[3] J. F. S. Gomes and F. R. Leta, "Applications of computer vision techniques in the agriculture and food industry: A review," *European Food Research and Technology*, vol. 235, no. 6, pp. 989–1000, 2012.

[4] E. S. Telles, D. A. C. Barone, and A. M. Da Silva, "Inteligência Artificial no Contexto da Indústria 4.0," pp. 130–136, 2020.

[5] S. Marshall, "Machine vision: Automated visual inspection and robot vision," *Automatica*, vol. 30, no. 4, pp. 731–732, 1994.

[6] K. Sage and S. Young, "Security applications of computer vision," *IEEE Aerospace and Electronic Systems Magazine*, vol. 14, no. 4, pp. 19–24, 1999.