

Implementation of FPGA accelerator architecture for Convolution Neural Network in Emotional Recognition System

Samson Immanuel J, Manoj G, Divya. P.S

Karunya Institute of Technology and Sciences, Coimbatore

Karunya Institute of Technology and Sciences, Coimbatore

Karunya Institute of Technology and Sciences, Coimbatore

Article History:Received:11 november 2020; Accepted: 27 December 2020; Published online: 05 April 2021

Abstract: The field of deep learning, artificial intelligence has arisen due to the later advancements in computerized innovation and the accessibility of data information, has exhibited its ability and adequacy in taking care of complex issues in learning that were not previously conceivable. The viability in emotional detection and acknowledging specific applications have demonstrated by Convolution neural networks (CNNs). In any case, concentrated Processor activities and memory transfer speed are necessitated that cause general CPUs to neglect to accomplish the ideal degrees of execution. Subsequently, to build the throughput of CNNs, equipment quickening agents utilizing General Processing Units (GPUs), Field Programmable Gate Array (FPGAs) and Application Specific Integrated circuits (ASICs) has been used. We feature the primary highlights utilized for productivity improving by various techniques for speeding up. Likewise, we offer rules to upgrade the utilization of FPGAs for the speeding up of CNNs. The proposed algorithm on to an FPGA platform and show that emotions recognition utterance duration 1.5s is identified in 1.75ms, while utilizing 75% of the resources. This further demonstrates the suitability of our approach for real-time applications on Emotional Recognition system.

Keywords: Adaptable Architectures, Convolutional Neural Networks (CNNs), Emotional Recognition System(ERS),Deep Learning, Dynamic Reconfiguration, Hardware Accelerator ,Field Programmable Gate Arrays (FPGAs).

Introduction

Lately, on account of the arrangement of enormous measures of valid information (Big Data: Audio, Video, Text, and so forth), and colossal advances inside the space of computerized material science innovations that offer huge processing power, there has been a recovery inside the space of figuring Artificial Intelligence (AI), strikingly inside the space of Deep learning (DL) (Y. Bengio et al. 2009), a subfield of Machine Learning (ML). The metric field limit unit arose in 2006 when an extended delays inside the space of neural organizations (NNs) examination (I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. 2016). A critical side in metric limit unit is that the organizations or potentially their loads aren't planned by men. All things being equal, they're gained from information utilizing a universally useful learning strategy (Rumelhart, David E and Hinton, Geoffrey E and Williams, Ronald J.1988). While utilizes calculations to investigate and gain from data, to make instructed choices, metric limit unit structures calculations in layers to make a Artificial neural Network (ANN) which will learn, and practically like human insight, will fabricate right determinations all alone (M. A. Nielsen.2015). In this way, instead of concocting calculations by hand, frameworks are regularly planned and prepared to execute thoughts in an incredibly way practically like what works out easily for people, and with precision for the most part Olympian executing in Human-Level (Mathworks.2018).

In Deep Learning, each one layer is intended toward discern choices on very surprising levels. A layer changes the delineation at the same level (beginning commencing PC record that perhaps pictures, sound, or text) to a illustration at the following, somewhat a ton of conceptual level (A. Krizhevsky, I. Sutskever, and G. E. Hinton.2012). Progressive layer collects the yield familiar articles to past layers, and the remaining layer identifies items. Even as we will in general set out through layers in great arrangement, the organization contributes an initiation that map speaks to a ton of and a ton of convoluted alternatives. The more profound you move into the organization; the channels start to be a great deal of mindful to a greater area of the recent trends. More elevated level layers intensify parts of the got inputs that square measure imperative for separation and smother unacceptable varieties.

EVOLUTION OF DL NETWORKS

CNN organizations zone unit consider joined the preeminent ground-breaking developments inside the field PC vision. The accomplishment of DL networks developed instantly recognizable quality during 2012 once

Krizhevsky et al. used CNNs in the direction successful in the yearly exercise of ImageNet, PC vision, (Y. LeCun, Y. Bengio, and G. Hinton.2015) Large Scale Vision Recognition Challenge (ILSVRC). AlexNet Model utilization, accomplished Associate in Nursing bewildering improvement on the grounds that the picture arrangement blunder conceived from twenty 6th (in 2011) to fifteen. ImageNet could be an ordinary benchmark dataset acclimated survey the exhibition of article identification, picture order calculations comprises millions of pictures variants appropriated more than a huge number of item classes.

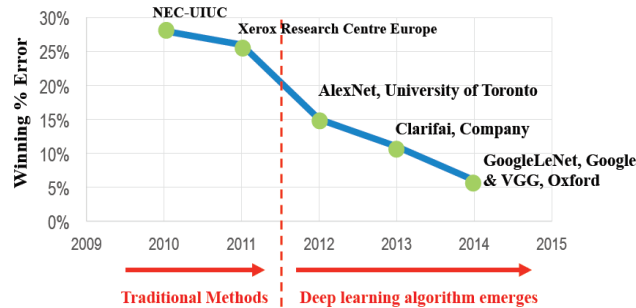


Figure:1. Results of ImageNet Competition (Image-Net.2018)

CNNs have accomplished significantly higher precision in grouping and shifted pc vision errands. The characterization exactness in ILSVRC improved to eighty eight.8%, 96.4%, and 93.3% (A. Deshpande.2018) within 2013 - 2015 rivalries, severally. Figure 1 show the precision misfortune used for the victor of ImageNet rivalries previously when there a rise of DL calculations. From that point, goliath has enterprises begun exploitation CNNs at their administrations. Microsoft, Google, Instagram, Amazon, Facebook, and Pinterest, and square measure by and by exploitation neural organizations for their image search, Bing's picture channels, programmed labeling calculations, item proposals, home channel personalization, and for their hunt foundation, severally (C. Zhang, D. Wu, J. Sun, G. Sun, G. Luo, J. Cong.2016). Be that as it may, the exemplary use- instance of CNNs intended for picture as well as discourse measure.

HARDWARE ACCELERATION OF DL

Toward hardware acceleration, the size of the CNN organization must exist amplified with adding together extra layers. Nonetheless, advancing extra and novel sort of neural network layers winds up in extra muddled CNN furthermore as high complexity models of CNN. Accordingly, billions no of tasks also a ton boundaries, furthermore like considerable registering assets expected to mentor and quantify the resultant large scale, CNN. Such necessities speak to a cycle dispute for General Purpose Processors (GPP). Subsequently, equipment quickening agents like FPGA, ASIC and GPU are utilized to support the yield of the CNN. In follow, CNNs prepared disconnected exploitation the backpropagation technique. At that point, the disconnected prepared CNNs won't to perform acknowledgment undertakings exploitation the feed-forward strategy. In this way, the outcomes and the another important parameter speed of the forward feeding techniques are viewed periodically,

CONVOLUTIONAL NEURAL NETWORKS (CNNs)

In this section, the primary tasks and wording engaged portrayed with the improvement of CNNs convolution, actuation capacities, standardization, qualities, and pooling of completely layers associated.

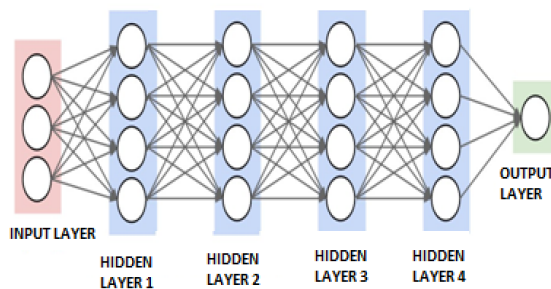


Figure :2. A regular Neural Network and ConvNet arrangements

CHALLENGES OF FPGA-BASED IMPLEMENTATION OF CNN

Usage of profound learning organizations, CNNs on FPGAs contains a scope of difficulties along with the need of a significant amount of capacity. Memory data measure and cycle assets on the request for billions of tasks for every second (V. Sze, Y.-H. Chen, J. Emer, A. Suleiman, and Z. Zhang.2018). AlexNet CNN has more than 60 million model boundaries requires memory of 250MB for putting away loads upheld thirty two cycle skimming point outline yet as necessities around one. Five billion tasks for every information from picture. Immense amount of capacity needed isn't upheld by the available modern FPGAs, thus the loads had the opportunity to be keep on outer memory and moved to the FPGA all through calculation. While not cautious usage of profound learning organizations and augmenting asset sharing, the execution probably won't chip away at FPGAs due to confined rationale assets.

The issue intensifies with extra convoluted models like VGG-CNN model with sixteen layers. Intra-yield correspondence parallelizes the calculation of one yield picture since it's the addition of n input-portion convolutions. Notwithstanding, between yield correspondence depends on figuring various yield FMs in equal. Likewise, convolutional layers square measure computational-driven though totally associated layers square measure memory focal.

Because innovation propels, Field Programmable Gate arrays actually fit in its size and capacities. Significant in order to claim a few systems for tending to the necessities for conservative executions of profound learning organizations. Tending to equipment asset limits needs utilize of methodology assets, and putting away of halfway prompts inside memories. Information move and strategy asset utilization are significantly reduced by the requesting of activities and decision of correspondence inside the usage of CNNs on FPGAs. Cautious programming of tasks may wind up in imperative decrease in outer activity and interior cushion sizes.

Proposed Method

The arranged CNN model is utilized for dissect the human inclination acknowledgment from the given dataset. There are few difficulties in CNN model plan like force, space and postponement. This investigation is focused on the usage of reconfigurable plan for CNN. It's applied to human inclination acknowledgment continuously application.

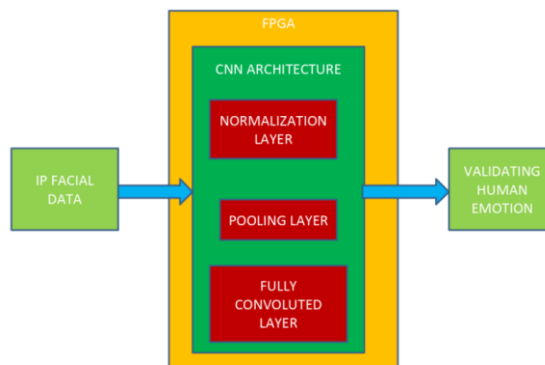


Figure:3. A regular Neural Network and ConvNet arrangements

The projected plan comprises of FPGA and a processing unit. CNN computations are performing through unique style of cycle element modules in FPGA. the most modules inside the process single meant square measure max-pooling, convolved muddled, data move, non-linearity, inclination move, also snake tree, which is demonstrated in Fig.3. The convolves muddled implied as traditional line buffer, as demonstrated in Fig. 3, to acknowledge convolution operations moreover on figure FC layer increase of matrix-vector. The pooling layer authorized inside the max pooling module is utilized to yield the most worth in the input information flows with a path of range a couple. The activation function of Convolution network is apply to the PC record stream of data misuse then on-linearity section. The viper hierarchy aggregates the halfway wholes produced while it convolves. At long last, information shift and predisposition move modules square measure chargeable for accomplishing dynamic quantization. The creators have widely used extraordinary stockpiling design of both pixel inputs and loads in on memory of off-chip prior to the acceleration strategy toward augment information use also minimize of electronic correspondence. The plan of CONV module is planned upheld the speeding up

strategies in anyway with an unmistakable association of Mack units as shown in Fig. 3. The Mack units of CONV module have been coordinated into Piy_ P of independent MAC blocks, with every Mack block contain Pix Mack unit to additionally limit the cradle peruse activities and consequently the partial sums developments. In addition, such association grants to handle variable (piece, step) sizes arrangements through generating totally various variations of CONV register exhibits during the accumulation.

LAYERS USED TO BUILD CONVNETS

As we tend to spoke to higher than, a regular ConvNet could be a grouping of layers, as well as each of a layer ConvNet changes 1 volume of initiations to an alternate throughout a differentiable perform. We will in general utilize 3 principle kinds of layers toward to make ConvNet structures: Convolutional Layer, Fully-Connected Layer and Pooling Layer (precisely the same as found in CNNs). These layers are stacked to make a entire ConvNet plan. The prototype of the ConvNet is given as below

Model Architecture:.

The extra subtleties beneath, anyway a clear ConvNet for CIFAR-10 order may have the plan [INPUT - CONV - RELU - POOL - FC]. in extra detail: INPUT [32x32x3] will hold the crude pixel estimations of the picture, for this situation a picture of width 32, stature 32, and with three shading channels R,G,B.

CONV layer (Y. Ma, M. Kim, Y. Cao, S. Vrudhula, and J.Seo. 2017) will register the yield of neurons that are associated with nearby areas in the info, each processing a speck item between their loads and a little district they are associated with in the information volume. RELU layer will apply an elementwise initiation work, POOL layer will play for spatial measurement (width, tallness) along with a down sampling activity, bringing about volume, for example, [16x16x12].

Fully connected layer force register the class scores, bringing about volume of size [1x1x12], where every one of the 10 numbers relate to a class score, for example, among the 10 classifications of CIFAR-10. Likewise with CNNs and as the name infers, every neuron in this layer will be associated with all the numbers in the past volume. ConvNet engineering is in the least difficult case a rundown of Layers that change the picture volume into a yield volume (for example holding the class scores).There are a couple of particular kinds of Layers. Each Layer acknowledges an info 3D volume and changes it to a yield 3D volume with a differentiable capacity

The Convolutional layer is the center structure square of a CNN network that wills a large portion of the technique work. Review and instinct without cerebrum stuff. How about we beginning examine what the CONV layer processes while not mind/neuron analogies. The CONV layer's boundaries contain an assortment of learnable channels. Each channel is small spatially (along measurement and stature), anyway reaches out through the total profundity of the info volume. for instance, an ordinary channel on an essential layer of a ConvNet may require size 5x5x3 (for example five pixels measurement and stature, three and three) because of pictures have profundity 3, the shading channels). all through the pass, we tend to slide convolve each channel transversely the measurement also stature of info volume and code spot item among the passages of channel and in this manner the contribution at any point. While we tend to go down the channel above the measurement, also tallness of the information volume we fabricate a 2-dimensional enactment map that offers the reactions of that channel at each special position. Naturally, the organization can learn channels that initiate once they see some sort of visual element like a hold of some direction or a smear of some tone on the essential layer, or in the long run whole wheel or honeycomb examples on top layers of the organization. Presently, we include a total arrangement of channels in every CONV layer (for example twelve channels), and everything about can fabricate a different 2D enactment map. We will load these enactment maps on the profundity measurement and production the yield volume.

POOLING LAYER

It is normal to irregularly embed a Pooling layer along Conv layers during a (Christopher Choy, JunYoung Gwak, and Silvio Savarese.2019) ConvNet plan. Its work is to progressively downsize the spatial size of the representation to scale back the amount of boundaries and calculation inside the organization, and consequently to conjointly the board overfitting. The Pooling Layer works severally on every profundity cut of the info and resizes it spatially, exploitation the goop activity represented in fig 4. The preeminent normal sort might be a pooling layer with channels of size a couple ofx2 functional with a step of two down samples each profundity cut inside the contribution by two on each measurement and also tallness, disposing of seventy fifth of the initiations. Every goop activity would during this case be taking a goop more than four numbers (minimal 2x2 area in some profundity cut). The profundity measurement stays unaltered. Extra normally, the pooling layer:

Pseudo code for Pooling Layer

- Obtain a volume of size $W1 \times H1 \times A1$ $W1 \times H1 \times A1$
- Essentially the two hyperparameters are taken into consideration for the portotype their spatial degree FF, the step SS,
- Produces a volume of size $W2 \times H2 \times A2$ $W2 \times H2 \times A2$ where:
- $W2 = (W1 - F) / S + 1$ $W2 = (W1 - F) / S + 1$
- $H2 = (H1 - F) / S + 1$ $H2 = (H1 - F) / S + 1$
- $A2 = A1$ $A2 = A1$

Introduces zero boundaries since it figures a fixed capacity of the info For Pooling layers, rarely to cushion the info utilizing zero-cushioning.

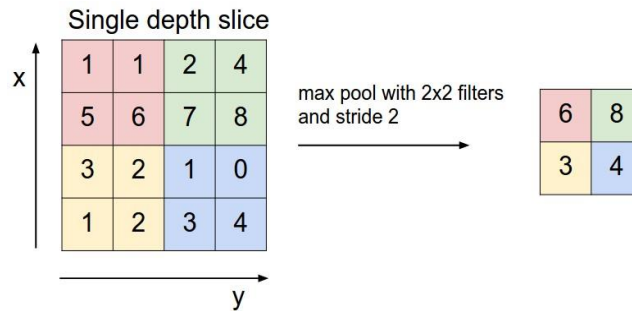


Figure:4. Little 2x2 square – max pooling

Fully-Connected Layer

In figure 5, the Fully Connected layer the Neurons inside a completely associated layer contain packed associations with each and every one or any enactments inside the past layer, like found in standard Networks. Their initiations will hence be processed by a network activity follow by an inclination counterbalance. See the Neural Network part of the notes for a great deal of information's

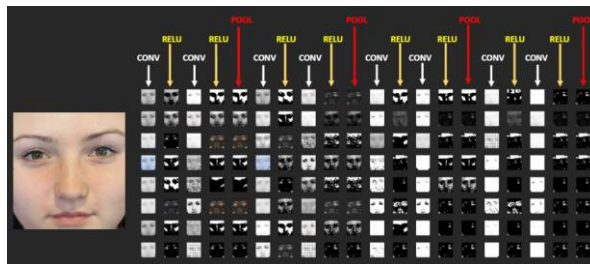


Figure:5. Simulation of Fully Connected Layer

Table. 1. Emotional Recognition Analysis

Correctness	Emotional Recognition						
	Neutral	Angry	Happy	Fearful	Sad	Boredom	Disgusted
Neutral	0.55	0	0	0	0.11	0.20	0
Angry	0	0.82	0	0	0	0	0
Happy	0.13	0.30	0.75	0.23	0	0	0.20
Fearful	0.19	0.2	0	0.85	0	0.11	0
Sad	0.12	0	0	0	0.86	0	0
Boredom	0.15	0	0	0	0.09	0.88	0
Disgusted	0	0	0	0	0	0	0.90

Converting FC Layers to CONV Layers

It is esteem noticing that solely qualification among Fully Connected layer and CONV layer be that the neurons inside the CONV layer square measure associated uniquely to a local district inside the information, which a few of the neurons during a CONV volume share boundaries. The neurons in each layers actually figure speck item, in this manner their valuable sort is indistinguishable. In this manner, it appears to be that it's capability to change over among

Fully Connected and CONV layers on behalf of partner level CONV layer there is a Fully Connected layer that actualizes a comparative advance work. The heap network would exist an outsized lattice that is chiefly zero beside at sure squares (because of local availability) any place the loads in a few of the squares square measure equivalent (because of boundary sharing). equally, any Fully Connected layer is brought back to life to a CONV layer.

FPGA accelerator implementation of the emotional recognition

The hardware implementation of the FPGA is done with the Xilinx XC7Z045 GPGA COT Board is shown in the Fig 6. The (Xilinx.2015) Xilinx Zynq™-7000 EPP ARM® dual-core Cortex™-A9 + 28 nm programmable logic board with the highly integrated TI XC7Z045 analog front end and RF components, enables software defined radio optimized for low power consumption for the implementation of the Emotional Recognition System(ERS).

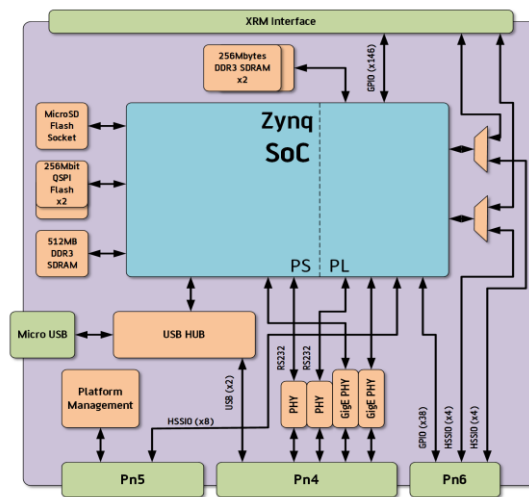


Figure: 6. Zynq XC7Z045 FPGA COT Board (Xilinx.2015)

A throughput-improved usage is given in official that also utilizes rationale segments along with DSP blocks, anyway the amount of multipliers keeps on being plentiful more modest than that of our RTL execution. Our norm and ascendible RTL execution of AlexNet beats the OpenCL style on an equal board with comparable FPGA usage (C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong.2015) by one.9X for the yield and furthermore the HLS style by > 2X for convolution layers. presents a pipelined quickening agent to plan totally different completely different } convolution layers onto diverse FPGA registering equipment to expand the asset usage. Notwithstanding, with the swelled scope of CNN layers, it gets more earnestly to with productivity allocate totally various assets to numerous layers while keeping the harmony between each pipeline stage. Generally, the RTL execution of CNN quickening agent gives significant execution benefit over elevated level combination based for the most part usage, that don't have reasonable equipment intensity. CNN RTL compiler with defined ascendible speeding up modules also allows quick pivot time that is appreciating significant level amalgamation techniques.

For assessment of our proposed approach, the uninhibitedly accessible German Emotion information base (EMO-DB) (Chen, M.; He, X.; Yang, J.; Zhang, H.2018) is utilized. It comprises of non-unconstrained, acted feelings by 10000 speakers, 6000 male and 4000 female. Every expression is distinguished by a solitary feeling having a place with one of seven classifications - neutral (N), happiness (H), anger (A), fear (F), sadness (S), boredom (B) or disgust (D). Only 493 utterances with a minimum of 80% human recognition accuracy and 60%

effortlessness are picked for our trials. Preparing is performed utilizing 70% of the absolute expressions and the calculation is assessed on the emotional expression.

This component is presented in Fig 7 where some metrics for evaluating the model. The consequence of arrangement will be in four potential cases, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In these four metrics, the Accuracy represents general information of the models performance.

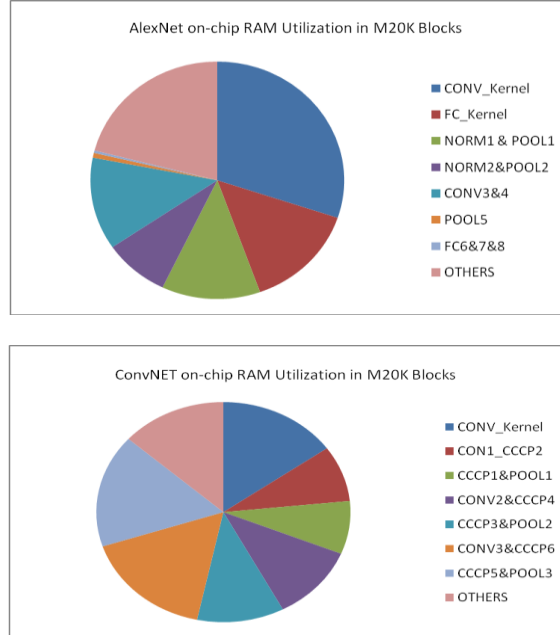


Figure:7. Utilization of FPGA Accelerator architecture

In any case, in peculiarity recognition issues, the quantities of glad examples are normally incredibly greater than the any remaining ones. Thou sly, if the model is basically set such that all of data sources are named glad, it can arrive at a fantastic precision. In this way, Precision and Recall are abused to defeat the deficiency of Accuracy eq. (3)

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

$$Fscore = \frac{2 * Recall * Precision}{Recall + Precision} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

This segment is introduced a few measurements for assessing the model. The result of classification will be in 4 possible cases, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In these four metrics, the Accuracy represents general information of the models performance. Notwithstanding, in abnormality location issues, the quantities of upbeat examples are regularly incredibly greater than the any remaining ones. Therefore, if the model is essentially set such that all of sources of info are delegated cheerful, it can arrive at a staggering precision. Accordingly, Precision and Recall are abused to conquer the weakness of Accuracy eq. (3).

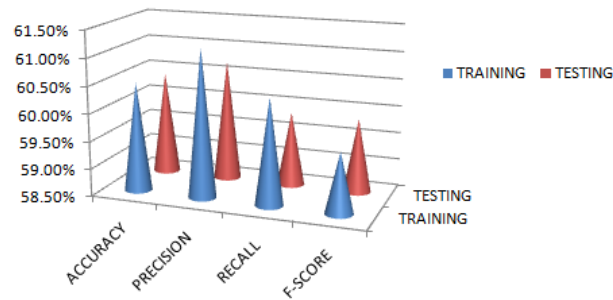


Figure: 8. Testing and Training segments for assessing the ConvNet MODEL

Conclusion

In this paper, ConvNet RTL compiler is proposed to quicken CNNs on FPGA stages, where the registering natives may be adequately ordered from the parameterized gear library. Agent CNN calculations of AlexNet and VGG have been shown on ZYNG FPGA board, which show a start to finish throughput of 114.5 GOPS and 117.3 GOPS, bringing about 1.9X improvement contrasted with an advanced plan on a similar FPGA board. We've additionally demonstrated all the more for the most part that more modest step and bigger quantities of highlights yield monotonically improving execution, which proposes that while more intricate calculations may have more noteworthy authentic force, basic yet quick calculations can be exceptionally serious. While confirming the basic finding that more features and dense extraction are useful, we have shown more importantly that these elements can, in fact, be as important as the supervised learning algorithm itself. The emotional recognition of the system is trained to the accuracy, precision, recall and F-Score of the real time system is estimated.

REFERENCES

- A. Krizhevsky, I. Sutskever, and G. E. Hinton.2012, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105.
- A. Deshpande.2018 "A Beginner's Guide To Understanding Convolutional Neural Networks [Online]". Available: <https://adeshpande3.github.io/ABeginner%27s-Guide-To-Understanding-Convolutional-NeuralNetworks/>.
- C. Zhang, D. Wu, J. Sun, G. Sun, G. Luo, J. Cong.2016. Energy-efficient CNN implementation on a deeply pipelined FPGA cluster, in: *ACM Int. Symp. On Low Power Electronics and Design (ISLPED)*, pp. 326–331.
- C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong.2015."Optimizing FPGA-based accelerator design for deep convolutional neural networks," in *Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*. ACM, pp. 161–170.
- Chen, M.; He, X.; Yang, J.; Zhang, H.2018. "3-D convolutional recurrent neural networks with attention model for speech emotion recognition." *IEEE Signal Process. Lett.* 25, 1440–1444.
- Christopher Choy, JunYoung Gwak, and Silvio Savarese.2019. "4d spatio-temporal convnets: Minkowski convolutional neural networks", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3075–3084.
- I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. 2016, *Deep learning*. MIT Press Cambridge.
- M. A. Nielsen.2015. "Neural networks and deep learning," *Determination Press*, USA.
- Mathworks.2018. "What Is Deep Learning? [Online]". Available: <https://www.mathworks.com/discovery/deep-learning.html/>.
- Rumelhart, David E and Hinton, Geoffrey E and Williams, Ronald J.1988, "Neurocomputing: Foundations of research," *ch. Learning Representations by Back-propagating Errors*, pp. 696–699.
- V. Sze, Y.-H. Chen, J. Emer, A. Suleiman, and Z. Zhang.2018. "Hardware for machine learning: Challenges and opportunities," in *Custom Integrated Circuits Conference (CICC)*, IEEE, 2018, pp. 1–8.
- Xilinx.2015. Zynq-7000 All Programmable SoC Overview, DS190(v1.8), Xilinx. https://www.xilinx.com/support/documentation/sw_manuals/x
- Y. Bengio et al.2009. "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127.
- Y. LeCun, Y. Bengio, and G. Hinton.2015, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436.

- Image-Net.2018. “The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [Online]”. Available: <http://image-net.org/challenges/LSVRC/>.
- Y. Ma, M. Kim, Y. Cao, S. Vrudhula, and J.Seo. 2017. “End-to-end scalable FPGA accelerator for deep residual networks,” in *Circuits and Systems (ISCAS), IEEE International Symposium on. IEEE*, pp. 1–4.

ABOUT THE AUTHORS

- Samson Immanuel J:** Assistant Professor, Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.
- Dr. Manoj G:** Assistant Professor, Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.
- Dr. Divya P.S:** Assistant Professor, Department of Mathematics, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.