1. J. Samson Immanuel [1], *2. G. Manoj [2], Shajin Prince [3], 4. H. Victor Du John [4]

[1,2,3,4] Assistant Professor, Karunya Institute of Technology and Sciences, Coimbatore, India.
ORCID: 1. 0000-0002-6914-7831; 2. 0000-0003-3543-5672; 3. 0000-0001-6708-6011; 4. 0000-0002-9889-2566

doi:10.15199/48.2021.07.26

Analysis of FPGA accelerator architecture for Fast Statistical Convolutional Neural Network in real time Emotional Recognition System

Abstract.. *Deep learning, an artificial intelligence area that emerged as a consequence of later developments in computerized innovation and the accessibility of data knowledge, has demonstrated its skill and adequacy in coping with complex learning problems that were previously unthinkable. (CNNs). Convolution neural network has shown the feasibility of emotional detection and acknowledging unique applications. In any case, concentrated processor activities and memory transfer speed are required, which causes general CPUs to fall short of achieving optimal execution levels. Following that, equipment quickening agents using General Processing Units (GPUs), Field Programmable Gate Array (FPGAs), and Application Specific Integrated Circuits (ASICs) were used to increase the throughput of CNNs. In addition, we include rules for improving the use of FPGAs for CNN speedup. The proposed algorithm is implemented on an FPGA platform, and results show that emotions regonition utterances of 1.25s are found in 1.85ms, consuming 85% of the resources. This illustrates the suitability of our approach for real-time Emotional Recognition device applications.*

*Streszczenie. Deep learning, dziedzina sztucznej inteligencji, która pojawiła się w wyniku późniejszych postępów w skomputeryzowanych innowacjach i dostępności wiedzy na temat danych, dowiodła swoich umiejętności i adekwatności w radzeniu sobie ze złożonymi problemami uczenia się, które wcześniej były nie do pomyślenia. Neuronowa sieć konwolucyjna wykazała wykonalność wykrywania emocji i rozpoznawania wyjątkowych zastosowań. W każdym razie wymagane są skoncentrowane działania procesora i szybkość transferu pamięci, co powoduje, że ogólne procesory nie osiągają optymalnych poziomów wykonania. W celu zwiększenia przepustowości CNN, zastosowano środki przyspieszające sprzętu, wykorzystujące jednostki przetwarzania ogólnego (GPU), programowalną macierz bramek (FPGA) i układy scalone specyficzne dla aplikacji (ASIC).. Proponowany algorytm jest zaimplementowany na platformie FPGA, a wyniki pokazują, że wypowiedzi regonacji emocji o długości 1,25s znajdują się w czasie 1,85 ms, co pochłania 85% zasobów. To ilustruje przydatność naszego podejścia do aplikacji urządzeń do rozpoznawania emocji w czasie rzeczywistym. (**Analiza architektury akceleratora FPGA dla szybkiej statystycznej sieci splotowej w systemie rozpoznawania emocji w czasie rzeczywistym**)*

Keywords: Adaptable Architectures, Fast statistical Convolutional Neural Networks (FSCNNs), Emotional Recognistion System(ERS),Deep Learning, Dynamic Reconfiguration, Hardware Accelerator ,Field Programmable Gate Arrays (FPGAs)
Słowa kluczowe: Sieci neuronowe, rozpoznawanie emocji, układy FPGA, Deep Lerning

Introduction

Recently, there has been a recovery in the space of figuring Artificial Intelligence (AI),especially in the space of Deep learning (DL) [1], a subfield of Ma, due to the arrangement of enormous quantities of valid information (Big Data: Audio, Video, Text, and so on), and colossal advances within the space of computerized material science developments that offer massive processing power.One of the most important aspects of the metric limit unit is that the organisations, and probably their loads, are not planned by men. They are gained from knowledge using a universally useful learning technique [3] if all other factors are equal. Although u uses calculations to investigate and learn from data in order to make informed decisions, the metric limit unit layers calculations to create an Artificial neural network.

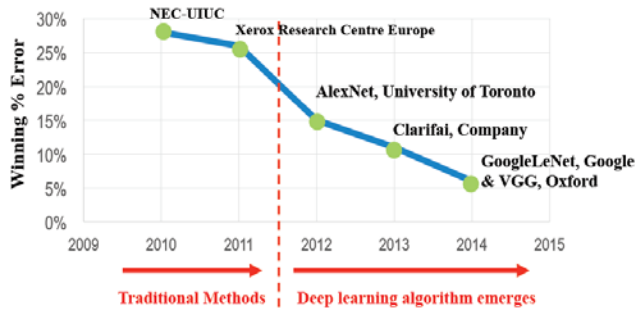

Fig. 1. Results of ImageNet Competition

Each layer in Deep Learning is built to distinguish choices on a surprising level is shown in the figure1 [8] Broad Scale Vision Recognition Challenge, Krizhevsky [7] et al. used CNNs in the direction of progress (ILSVRC). Since the picture arrangement blunder conceived from twenty sixth (in 2011) to fifteenth (in 2015), AlexNet Model use has made a befuddling progress. ImageNet may be a standard benchmark dataset, according to others.

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% [9]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 on, goliath has started using CNNs in their administrations. Adobe, Google, Instagram, Amazon, Facebook, and Pinterest, to name a few, use neural networks for image search, Bing's picture channels, programmed labeling estimates, object proposals, home channel personalization, and hunt base, among other things [10].

Fast convolutional neural networks (FCNNS)

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 is shown in the figure 2.

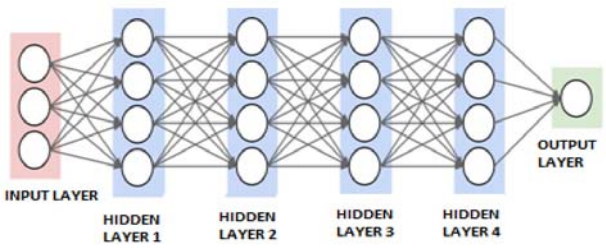

Fig. 2 Fast Neural Network and ConvNet arrangements

Usage Usage of profound learning organizations, FSCNNs 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 [11]. AlexNet CNN has more than 60 million model boundaries requiresmemory of 250MB for putting away loads upheld thirty two cycle skimming point outline yet as necessities around one.

Proposed Fast SCNN

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

The projected plan comprises of FPGA and a processing unit. FSCNN computations are perform through unique style of cycle element modules in FPGA. the most modules inside the process single ment square measure max-pooling, convolved mouddled, data move, non-linearity, inclination move, also snake tree, which is demonstrated in Fig.4.The convolver muddled implied as traditional line buffer, as demonstrated in Fig. 4, to acknowledge convolution operations moreover on figure FC layer increase of matrix-vector.

The prototype of the Fast 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 [12] 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 downsampling 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

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.

Introduces zero boundaries since it figures a fixed capacity of the infoFor Pooling layers, rarely to cushion the info utilizing zero-cushioning is shown in the figure 3.

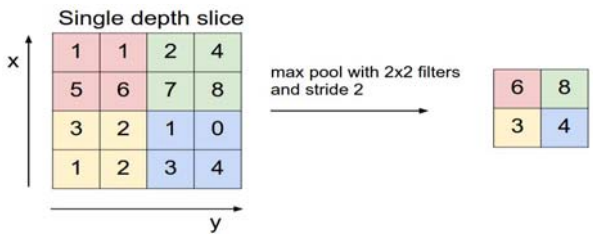

Fig. 3. little 2x2 square – max pooling

Fully-Connected Layer

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 FER dataset is shown in the table 1.

Table . 1. Emotional Recognization System parameters

Correct	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.

Analysis of FPGA accelerator implementation of the emotional recognition

The hardware implementation of the FPGA is done with the Xilinx XC7Z045 GPGA COT Board. The [13] hardware implementation of the Emotional Recognition System (ERS) is shown in the figure 4 which consist of the three layers the ROM , Covariance Calculation and Eigen Calculation. The video image is then processed with the Euclidean distance and the emotional recognition is done.

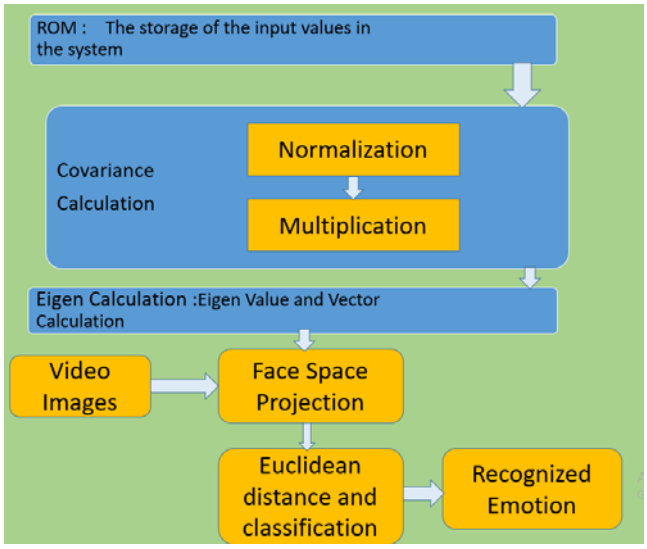

Fig 4: Hardware Implementation of Emotional Recognition

For assessment of our proposed approach, the uninhibitedly accessible German Emotion information base (EMO-DB) [15] 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 is shown in the figure 5 .

This component is presented some metrics for evaluating the model. The consequence of arrangement will be in four potential cases, namely True Positive (TP), True Negative (TN), for the FPGA utiisation is shown in the Table 2.

Table 2:FPGA Acceertor Utilisation

Module	DSP48E	BRAM (18K)	Flip Flop	LUT
ADDLAST_1U_S	0	0	1150	440
APPEND ZEROS	0	0	3326	780
CONV2D_XIL_NOP	41	32	4529	10121
CONV2D_XIL_NOP	32	21	3544	7666
CONV2D_XIL_NOP	33	21	4122	10121
CONTROL_S_X	25	42	3244	14232
EXTRACTPIXES	0	0	250	130
POOL2D_NOP	5	6	3166	1088
POOL2D_NOP	6	5	1195	588
TRUNCATEDWIDTH	0	0	1191	322
TOTAL REQUIRED	122	155	23224	35343
TOTAL AVAIABLE	150	156	10999	53343
UTILISATION	85%	62%	13.2%	69%

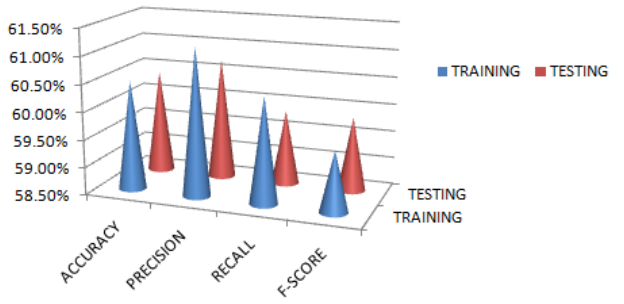

Fig. 5.Testing and Training segements for assessing the FSCNN MODEL

Conclusion

In this paper, ConvNet RTL compiler is proposed to quicken CNNs on FPGA st`ages, 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 XC7Z020-1CLG484C 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 recognisation of the system is tranined to the accuracy,precision,recall and F-Score of the real time system is estimated.

Authors

Mr. J. Samson Immanuel completed his B.E from Periyar University and M.E degree from Karunya University, India. Currently he is working as Assistant Professor of ECE in karunya institute of technology and sciences and doing his part time research.

Dr. Manoj G received his B. Tech degree from Amritha University, Coimbatore and M.Tech degree from Karunya University, Coimbatore, India. Corresponding Author: manojpillai@karunya.edu

Dr. Shajin Prince completed his B.Tech and M.Tech degree from Karunya University, India. Currently he is working as Assistant Professor of ECE in karunya institute of technology and sciences.

Dr. H. Victor Du John completed his B.E from Anna University Chennai, M.Tech and Ph.D degree from Karunya University, India. Currently he is working as Assistant Professor of ECE in karunya institute of technology and sciences.

REFERENCES

[1] Y. Bengio et al., “Learning deep architectures for ai,” Foundations and trends® in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009.

[2] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning. MIT Press Cambridge, 2016.

[3] Rumelhart, David E and Hinton, Geoffrey E and Williams, Ronald J, “Neurocomputing: Foundations of research,” ch. Learning Representations by Back-propagating Errors, pp. 696–699, 1988.

[4] M. A. Nielsen, Neural networks and deep learning. Determination Press, USA, 2015.

[5] Mathworks. What Is Deep Learning? [Online]. Available: https://www. mathworks.com/discovery/deep-learning.html/, 2018.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[7] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, p. 436, 2015.

[8] A. Deshpande. A Beginner’s Guide To Understanding Convolutional Neural Networks [Online]. Available: https://adeshpande3.github.io/ABeginner%27s-Guide-To-Understanding-Convolutional-NeuralNetworks/, 2018.

[9] C. Zhang, D. Wu, J. Sun, G. Sun, G. Luo, J. Cong, Energy-efficient CNN implementation on a deeply pipelined FPGA cluster, in: ACM Int. Symp. On Low Power Electronics and Design (ISLPED), 2016, pp. 326–331.

[10] V. Sze, Y.-H. Chen, J. Emer, A. Suleiman, and Z. Zhang, “Hardware for machine learning: Challenges and opportunities,” in Custom Integrated Circuits Conference (CICC), 2018 IEEE. IEEE, 2018, pp. 1–8.

[11] Y. Ma, M. Kim, Y. Cao, S. Vrudhula, and J.-s.Seo, “End-to-end scalable FPGA accelerator for deep residual networks,” in Circuits and Systems (ISCAS), 2017 IEEE International Symposium on. IEEE, 2017, pp. 1–4.

[12] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3075–3084, 2019.

[13] Xilinx, *Zynq-7000 All Programmable SoC Overview*, DS190(v1.8), Xilinx, 2015.

[14] C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, “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, 2015, pp. 161–170.

[15] Chen, M.; He, X.; Yang, J.; Zhang, H. 3-D convolutional recurrent neural networks with attention model for speech emotion recognition. *IEEE Signal Process. Lett.* **2018**, *25*, 1440–1444.