International Journal of Future Generation Communication and Networking
Vol. 13, No. 1, (2020), pp. 203-210

A MACHINE LEARNING BASED ARCHITECTURE IMPLEMENTATION FOR OBJECT DETECTION

S.Mohanraj[1], S.Guru Prasad[2], S.Gurumoorthy[2], C.Karthik[2]
Assistant Professor[1], UG Scholar[2]
Department of Electronics and Communication Engineering,
M.Kumarasamy College of Engineering, Karur, Tamil Nadu, India.

Abstract

The ConvNet (CNN) is a neural network technique that has been extensively employed in detect and recognize object, video analytics and image classification. ConvNet based algorithms aren't easy to be deployed in the embedded devices as the ConvNet is memory expensive and compute-intensive. There are many studies that have reviewed the implementation of ConvNet based classification models on hardware, a vast area of implementation ConvNet on FPGA is yet to be explored. This work proposed the half-precision floating point ConvNet implementation to detect object with Tiny YOLO v2 on Intel Altera Cyclone-IV FPGA using Verilog HDL. In view of the resource constrains of an FPGA in relationswith memory bandwidth, computational resources along with on-chip memory, the pre-processing of data is planned so that the concatenationof batch normalization addicted to convolution layer.

2. INTRODUCTION

YOLO is a solitaryConvNet that predicts the multiple object alongside with the box's class probability. The full image is trained and the detection process is optimized by the YOLO. This technique has numerous benefits than the traditional methods (RCNN, Fast RCNN) of the object detection.YOLO works extremely faster than the previous methods of detecting the object. Though the detection is considered as a regression process, it doesn't require any complex pipeline process.YOLO archives twice the mean average (mAP) precision than the other system. The image is globally reasoned by YOLO while predicting objects of the given input. Similar to regional proposal and sliding window techniques, YOLO also considers the entire image while training and testing the image hence the image classes contextual information and their appearance are encoded.The natural images are trained and tested in the network, the various detection methods such as DPM, R-CNN areoutperformed by YOLO. YOLO is generalizable and is less possible to break when employed in new domains. Here in our proposed work we use the Tiny YOLO v2.

3. LITERATURE SURVEY

Caiwen Ding et al., presented a work to prevail the substantial accuracy loss, decreased utilization of DSP blocks in FPGAs. The proposed architecture is REQ-YOLO which is the basis for detecting objects that is aware of resources and weight quantization which includes the method of block-circulant matrix and a diversified weight quantization employing ADMM[1]. Ren et al., introduced a RP technique for the object detection framework facilitating image convolution features through Regional Proposal Network (RPN).The convolutional features can be shared by training Fast R-CNN and alternating optimization RPN. High class Region proposals (RP) is generated with the training of Regional Proposal networks (RPN) that is employed with Fast Region-based CNN in the object detection process [2].

ISSN: 2233-7857 IJFGCN

Wei.L et al., proposed a technique for the detecting the objects in the input images with the insight of single DNN. The method known as Single-Shot MultiBox Detection discretizes bounding boxes output space into a default boxes based on various aspect ratios including the feature map measure per location[3].Lin Bai et al., introduced a scalable, high performance accelerator using ConvNet deployed on FPGA employing the technique of Depth wise Separable Convolution. The CNNs including MobileNet Version 2, Xception employs the embedded systems with Depth wise distinguishable convolution which in turn reduces the operations and parameters with reduced loss in accuracy. The proposed technique provides the balance between processing speed and resources of the hardware, the accelerator fits different size FPGAs[4].

4. EXISTING SYSTEM

The prevailing system that proposes a hardware computing architecture to identify face that classify the given input image as human face or not. This digital architecture was designed and verified in MATLAB Simulink using Xilinx block set. This was also released in a FPGA platform.Learned filter were deployed to classify based fully on features than the assumptions used by the image processing methods.

4.1 Digital Hardware Architecture

In hardware realization of this proposal, the architecture functions as a multi-stage filtering, decimation, sub-sampling and non-linear ConvNet decision making algorithm. There are two stages in this realization.

In the initial stage, each and every frame of the input video is filtered and sub-sampled with a factor of two in all the dimensions. It is then convolved by four different linear FIR kernels with a mask size 5x5. The obtained filtered output images from FIR kernels are scaled, biased and further biased with dissimilar values represented as $b_{1,k}$, α_k and $b_{1,2,k}$ where k = 1, 2, 3,4. Four sets of intermediate output images are provided by scaling, filtering, biasing and non-linear decision blocks(DEC). The above operations are applied for all frames in the video. The decision regions from this stage form sub-images which are then applied to the succeeding stage of CovnNet algorithm.

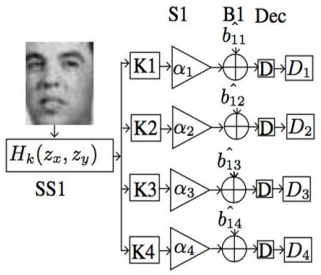


Fig. 4.1 First stage

In the next stage, sub-image outputs of the starting stage are given as input to this stage. This stage is further sub-divided into two architecture. In first architecture, the first quad images are sub-sampled and further convoluted with 8 diverse kernels represented as M1, M2, M3…, and M8 with a kernel of size 3x3. The operations performed in first stage is again implemented in this step of this stage. While the second algorithm pairs with output of former stage in an order denoted by (P_k, Q_k) and using 6 FIR filter kernels, sub-sampling and convolution is performed again. These pairs are separately summed and scaled, biased with distinct values represented as $b_{2,p}$, β_p and $b_{2,2,p}$ where p = 9,10, …,14. The resulting output is processed over DEC. The first architecture produces 8 images

ISSN: 2233-7857 IJFGCN

while the second architecture produces 6 output images. Thus, overall 14 images areattained as output.

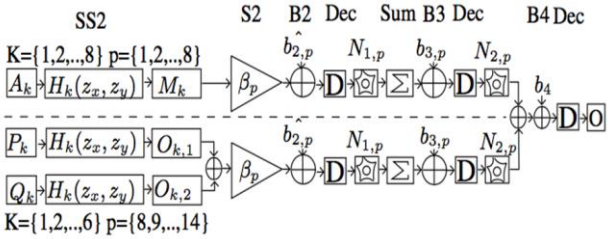


Fig. 4.2 Second stage

These images of size 6x7 are vectorized and a classical dot product operation is undergone with its respective neuron block(6x7) represented as $N_{1,p}$ where p=1,2,3,4,…,14 providing 14 different output images of same size. Weight metric is calculated and 14 dissimilar values are obtained. Now these are again biased and then perform dot product function with the next or second layer of neural network represented as $N_{2,p}$. Then obtained values are summed, biased (b4) and sent through the final DEC to attain the desired output. From this obtained value, the segregation of a given sample image as face or not is done. If the result is positive, then the given sample image is a face whereas the negative result indicates that the given sample image is not a face.

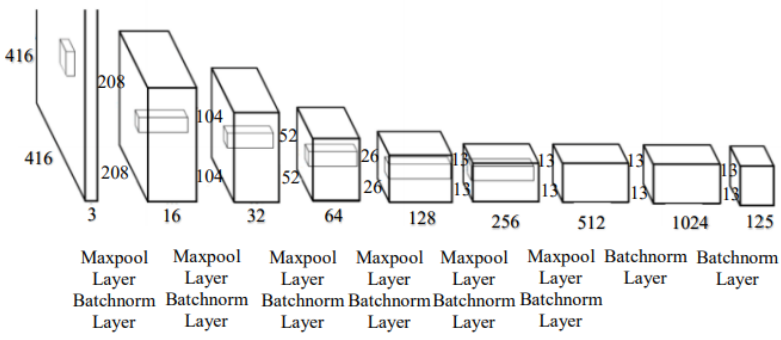
Implementing the face detection using MATLAB Simulink with 1164 test images, the implementation provided an accuracy of 96.56% where 1124 images where correctly detected. While realizing this in a FPGA provided accuracy above 90%. This proposed work could be a fair candidate in different face recognition and face detection tasks.

5. PROPOSED SYSTEM

In our proposed method, we desire to implement Tiny YOLO v2 in Intel FPGA to detect object using a Half-precision (16-bit) floating point. It is realized using Verilog HDL.

5.1 Tiny YOLO v2 Architecture

YOLO possess a unique approach that apply a single ConvNet on the whole image and identifies the multiple objects with boxes bounding it. Whereas the previous methodologies repurpose classifiers to detect object information in the frame or image. The Tiny YOLO v2 consists of nine convolutional layers, where each layer is equipped by means of a ReLU activation function in additionwith batch normalization operation strew with 6 max pooling layers and region layer. The resolution of the given input image to the Tiny YOLO v2 is 416 × 416 pixel.



ISSN: 2233-7857 IJFGCN

Fig. 5.1 Architecture of Tiny YOLO v2

5.2 Convolution Layer

The feedforward process is employed to detect object in Tiny YOLO v2. In feed forward computation period, convolution layer itself occupies over 90%. So, in Tiny YOLO v2,convolution layer is reduced to progress the accelerator performance. For a solitary input image, billions of additions and multiplication operations are performed between filters and local regions in convolution layer.

$$X^{(i)} = \sum_{i=1}^{n} (x^{(n)} \times w^{(k)}) + b$$

where,$X^{(i)}$is theOutput feature Pixel, $x^{(j)}$ is theInput feature Pixel, $w^{(k)}$represents Convolution weights and b is the Convolution bias.

The overallsum of operations performed by the convolution layer can be calculated using the following equation.

$$\text{Operations} = 2 \times N_{in} \times K \times K \times N_{out} \times H_{out} \times W_{out}$$

where,N_{in} is the No. of input channels, K representing Filter size, N_{out} is count of filters, $H_{out}$is theOutput feature Height and W_{out} = Output feature Width

The above equation disregards the determination of the amount of operations of batch normalization processing and ReLU activation at every layer. Space complexity describes the memory requirement. Weight is the important parameter in Tiny YOLO v2 in convolution layer. The total sum of weights in convolutional layer represented as

$$\text{Weights} = N_{in} \times K \times K \times N_{out}$$

5.3 Activation Function

The need of activation function in CNN architecture is to perform a transformation process on the given input before it proceeds to the pooling layer. The most frequently used activation functions of CNN are Sigmoidal activation function. This activation function is bounded by minimum and maximum value and thus initiating saturation neuron in the top layers of the network. Leaky ReLU can likewise be used and it may cause the change in weight that makes the functions to never activate again in any data point.

5.4 Pooling layer

Dimensional reduction of the given input in ConvNet can be done using the Pooling layer. The pooling layer removes the redundant information and retains the most critical information. Average pooling and maximum pooling are the characteristic pooling functions. The Max pooling function returns the highest value and the average function returns the average value. The equations of the pooling function are

$$S[i, j] = \max\{S[i;, j']: i \leq i' < i+, j \leq j' < j + p\}$$

$$S[i, j] = \text{average}\{S[i;, j']: i \leq i' < i+, j \leq j' < j+ p\}$$

ISSN: 2233-7857 IJFGCN

5.5 Batch Normalization

The latter layer in network is batch normalization which provide unit variance or zero mean for input at any layer of Tiny YOLO v2. The normalization function is done to the preceding output by subtracting and dividing the batch mean and batch variance respectively that has been obtainedfrom the convolution layer. After this operation,output isshifted, scaled. During the CNN training stage,values of the variance, scale, bias and means are generated. These obtained values make each layer to learn independently and the overfitting is reduced due to the regularization effect. The batch normalization equationis represented as

$$X^{(J)} = \frac{(X^{(I)} - \mu)}{\sqrt{\sigma^2 + \xi}}$$

where, $X^{(J)}$is Output Pixel after batch normalization, $X^{(I)}$represents Output Pixel after convolution, μ is the Mean value, $\sigma^2$is the Variance, ξrepresents the Constant value.

5.6 Block Diagram

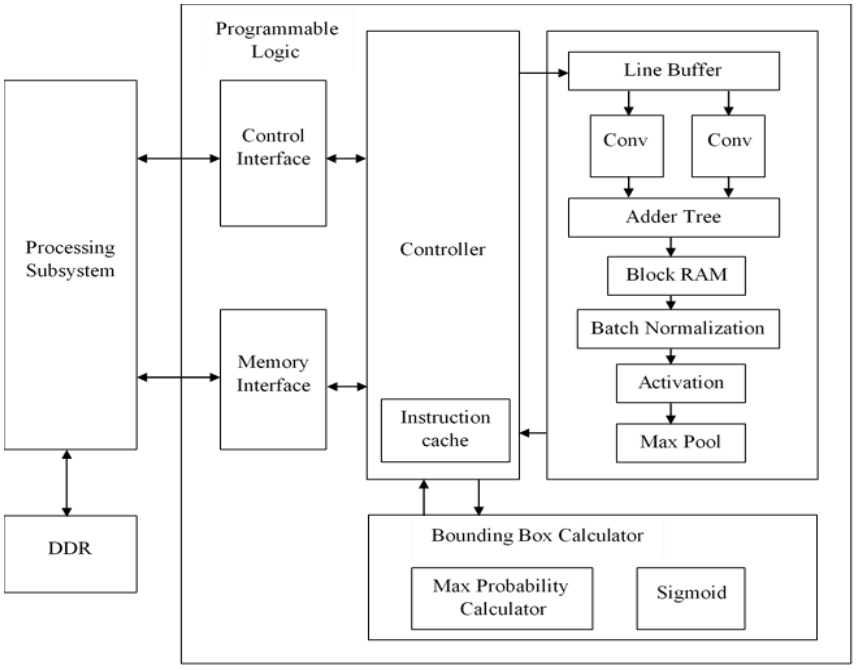


Fig. 5.3 Block diagram of Proposed work

As already discussed, the algorithms buffer the given input to the ConvNet. There are multiple convolution layers that process the input from the buffer and the output from this convolution layer are combined by adder tree which is followed by convolution layer. Then these are forwarded to the block RAM for Batch Normalization process where it provides the unit and zero variance. The shifted and scaled output from the preceding process is pre-transformed by activation function, then pooled by means of the Max Pooling operation in the pooling layer.

The bounding box calculator which mainly focuses only on bounding the detected object in the given image after processing through the CNN. Using the sigmoid function and the max

ISSN: 2233-7857 IJFGCN

probability calculator, the bounding of the recognized object in image is executed and thus object recognized in the given image.

6. RESULT

Initially Tiny YOLO v2 were implemented directly with the floating-point precision on Intel Altera Cyclone-IV. The utilization of hardware resource exceeded the hardware limitation as in utilize over 150% of RAM, 170% of logic and 70% of the DSP blocks in the hardware. Then to implement and outwork it, first the pre-processing batch normalization technique is approached which managed to reduce the exceeded hardware utilization by the algorithm. Instead of floating-point precision, half-precision floating point which is 16-bit is employed for further enhancement of algorithm to work efficiently.

Fig. 6.1 Predicted Image (Test Image) by Intel Altera Cyclone-IV using half-precision floating point Tiny YOLO v2

The image in Fig. 6.1 is the predicted test image of the half-precision floating point Tiny YOLO v2 on Intel Altera Cyclone IV. The hardware predicted the objects on the image with high accuracy.

7. CONCLUSION

The existing system that predicts whether the given image is a face or not were the input given of resolution 32x36 and of the format PGM (Portable Grey Map). The most frequently used image format like JPEG cannot be processed and also the real-time high-resolution images cannot be cast-off using the existing system. Our proposed system overcame both these limitations i.e. JPEG format and higher resolution images can be given as input to FPGA hardware for recognizing object with enhanced and optimized algorithm using half-precision floating point Tiny YOLO v2 in addition with batch normalization pre-processing. The proposed work is realized in the Intel Altera Cyclone-IV and obtained the result with high accuracy.

REFERENCE

1. Caiwen Ding, Shuo Wang, Ning Liu, Kaidi Xu, Yanzhi Wang, Yun Liang, “REQ-YOLO: A Resource-Aware, Efficient Quantization Framework for Object Detection on FPGAs”, pp. 33-42, 2019.
2. Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. “Faster RCNN: Towards Real-Time Object Detection with Region Proposal Networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1137-1149, 2016.

ISSN: 2233-7857 IJFGCN

3. T.Abirami, Dr.S.Palanivel Rajan, “Detection of poly cystic ovarian syndrome (PCOS) using follicle recognition techniques”, Bioscience Biotechnology Research Communications, ISSN: 0974-6455, Vol. 12, Issue : 01, pp. 1-4, DOI: 10.21786/bbrc/12.1/19, 2019.
4. Dr.S.Palanivel Rajan, “Enrichment of ECG Quality using Independent Component Analysis for Dynamic Scenario by Eliminating EMG Artifacts”, Advances and Applications in Mathematical Sciences, ISSN No.: 0974-6803, Vol. No.: 18, Issue : 2, pp. 219-237, 2018.
5. Dr.S.Palanivel Rajan, S.Suganya, “Design of Loop Antenna for the Human Brain Signal Analysis”, Indian Journal of Science and Technology, Online ISSN No.: 0974-5645, Print ISSN No.: 0974-6846, Vol. No.: 11, Issue: 10, pp. 1-6, DOI: 10.17485/ijst/2018/v11i10/120829, 2018.
6. M.Paranthaman, Dr.S.Palanivel Rajan, “Design of E and U Shaped Slot for ISM Band Application”, Indian Journal of Science and Technology, Online ISSN No.: 0974-5645, Print ISSN No.: 0974-6846, Vol.: 11, Issue: 18, pp. 1-3, DOI: 10.17485/ijst/2018/v11i18/123042 2018.
7. C.Vivek, S.Palanivel Rajan, “Z-TCAM : An Efficient Memory Architecture Based TCAM”, Asian Journal of Information Technology, ISSN No.: 1682-3915, Vol. No.: 15, Issue : 3, pp. 448-454, DOI: 10.3923/ajit.2016.448.454, 2016.
8. S.Vijayprasath, R.Sukanesh, S.Palanivel Rajan, “Assessment of relationship between heart rate variability and drowsiness of post operative patients in driving conditions”, JoKULL Journal, ISSN No.: 0449-0576, Vol. 63, Issue 11, pp. 107 – 121, 2013.
9. Paranthaman, M., and S. Palanivel Rajan. "Design of Triple C shaped Slot Antenna for Implantable Gadgets." Current Trends In Biomedical Communication And Tele–Medicine (2018): 40. DOI: 10.21786/bbrc/11.2/6
10. S.Palanivel Rajan, R.Sukanesh, S.Vijayprasath, “Design and Development of Mobile Based Smart Tele-Health Care System for Remote Patients”, European Journal of Scientific Research, ISSN No.: 1450-216X/1450-202X, Vol. No. 70, Issue 1, pp. 148-158, 2012.
11. M. Paranthaman, "T-shape polarization reconfigurable patch antenna for cognitive radio," 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM), Chennai, 2017, pp. 927-929. doi: 10.1109/ICONSTEM.2017.8261338
12. S.Palanivel Rajan, R.Sukanesh, S.Vijayprasath, “Analysis and Effective Implementation of Mobile Based Tele-Alert System for Enhancing Remote Health-Care Scenario”, HealthMED Journal, ISSN No. : 1840-2291, Vol. No. 6, Issue No. 7, pp. 2370–2377, 2012.
13. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, “SSD: Single Shot MultiBox Detector”, Computer Vision – ECCV, pp. 21–37, 2016.
14. Lin Bai, Yiming Zhao, Xinming Huang, “A CNN Accelerator on FPGA Using Depth wise Separable Convolution”, IEEE Transactions on Circuits and Systems-II, pp. 1415-1419, 2018.
15. S.Palanivel Rajan, T.Dinesh, “Statistical Investigation of EEG Based Abnormal Fatigue Detection using LabVIEW”, ”, International Journal of Applied Engineering Research, Vol. 10, Issue 43, pp. 30426-30431, 2015.
16. L. RAMESH, T.ABIRAMI, “Segmentation of Liver Images Based on Optimization Method”, International Journal of Pure and Applied Mathematics, Online ISSN No.: 1314-3395, Print ISSN No.: 1311-8080, Vol. No.: 118, Issue No.: 8, pp. 401-405, 2018.
17. S.Palanivel Rajan, V.Kavitha, “Diagnosis of Cardiovascular Diseases using Retinal Images through Vessel Segmentation Graph”, Current Medical Imaging Reviews Online ISSN No.:

ISSN: 2233-7857 IJFGCN

1875-6603, Print ISSN No.: 1573-4056, Vol. No.: 13, Issue : 4, pp. 454-459, DOI : 10.2174/1573405613666170111153207, 2017.

18. S Mohanapriya, M Vadivel, "Automatic retrival of MRI brain image using multiqueries system", 2013 International Conference on Information Communication and Embedded Systems, INSPEC Accession Number: 13485254, Electronic ISBN: 978-1-4673-5788-3, DOI: 10.1109/ICICES.2013.6508214, pp. 1099-1103, 2013.
19. S.Palanivel Rajan, "Review and Investigations on Future Research Directions of Mobile Based Tele care System for Cardiac Surveillance", Journal of Applied Research and Technology, Vol.13, Issue 4, pp.454-460, 2015.
20. S.Palanivel Rajan, R.Sukanesh, "Experimental Studies on Intelligent, Wearable and Automated Wireless Mobile Tele-Alert System for Continuous Cardiac Surveillance", Journal of Applied Research and Technology, ISSN No.: 1665–6423, Vol. No. 11, Issue No.: 1, pp.133-143, 2013
21. S.Palanivel Rajan, R.Sukanesh, "Viable Investigations and Real Time Recitation of Enhanced ECG Based Cardiac Tele-Monitoring System for Home-Care Applications: A Systematic Evaluation", Telemedicine and e-Health Journal, ISSN: 1530-5627, Online ISSN: 1556-3669, Vol. No.: 19, Issue No.: 4, pp. 278-286, 2013.
22. M.Paranthaman, S.Palanivel Rajan, "Design of H Shaped Patch Antenna for Biomedical Devices", International Journal of Recent Technology and Engineering, ISSN : 2277-3878, Vol. No. 7, Issue:6S4, pp. 540-542, Retrieval No.: F11120476S4/19©BEIESP, 2019.
23. S.Palanivel Rajan, et.al., "Intelligent Wireless Mobile Patient Monitoring System", IEEE Digital Library Xplore, ISBN No. 978-1-4244-7769-2, INSPEC Accession Number: 11745297, IEEE Catalog Number: CFP1044K-ART, pp. 540-543, 2010.
24. S.Palanivel Rajan, et.al., "Cellular Phone based Biomedical System for Health Care", IEEE Digital Library Xplore, ISBN No. 978-1-4244-7769-2, INSPEC Accession Number: 11745436, IEEE Catalog Number: CFP1044K-ART, pp.550-553, 2010.
25. Rajan, S., & Paranthaman, M. (2019). Characterization of compact and efficient patch antenna with single inset feeding technique for wireless applications. Journal of Applied Research and Technology, 17(4).
26. Dr.S.Palanivel Rajan, L.Kavitha, "Automated retinal imaging system for detecting cardiac abnormalities using cup to disc ratio", Indian Journal of Public Health Research & Development, Print ISSN: 0976-0245, Online ISSN: 0976-5506, Vol. No.: 10, Issue : 2, pp.1019-1024, DOI : 10.5958/0976-5506.2019.00430.3, 2019.
27. T.Abirami, S.Palanivel Rajan, "Cataloguing and Diagnosis of WBC'S in Microscopic Blood SMEAR", International Journal of Advanced Science and Technology, P-ISSN: 2005-4238, E-ISSN: 2207-6360, Vol. 28, Issue No. 17, pp. 69-76, 2019.
28. Rajan S. P, Paranthaman M. Novel Method for the Segregation of Heart Sounds from Lung Sounds to Extrapolate the Breathing Syndrome. Biosc.Biotech.Res.Comm. 2019;12(4).DOI: 10.21786/bbrc/12.4/1, 2019.
29. Dr.S.Palanivel Rajan, "Design of Microstrip Patch Antenna for Wireless Application using High Performance FR4 Substrate", Advances and Applications in Mathematical Sciences, ISSN No.: 0974-6803, Vol. No.: 18, Issue : 9, pp. 819-837, 2019.
30. M Paranthaman, G.Shanmugavadivel "Design of Frequency Reconfigurable E-Shaped Patch Antenna for Cognitive Radio" International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.20 (2015) pp.16546-16548

ISSN: 2233-7857 IJFGCN