

Deep Learning based Object Tracking in Field Hockey using FPGA

M. Praveena¹, V. Vinoth Thyagarajan², V.R. Venkatasubramani³, S. Rajaram⁴

Post graduate Student, Department of Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, India¹

Assistant Professor, Department of Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, India^{2,3}

Professor and Head, Department of Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, India⁴

Abstract— In VLSI, several transistors are integrated onto a single microchip. Enhancing or extracting information from images is known as image processing. Image processing algorithms may be quickly and adaptably implemented because of VLSI technology, particularly FPGAs as it is reconfigurable and do functions in parallel. Deep learning is a type of machine learning that makes use of multi-layered neural networks to learn from vast amounts of data. In hockey, deep learning algorithms like CNN improve player performance. Defending players, team tactics, ball speed, player movements, field and weather conditions, and game dynamics are all variables that affect performance. Deep learning algorithms are used to analyze player strategy, violence detection, missing person searches, item and weapon detection, replays, real-time videos, scores, and match reconstruction. The overall idea of our work's experimental setup consists of a camera connected to an SD card and a VGA interface with an FPGA. Using the SD card interface, the SD-card talks to the FPGA. The FPGA reads and stores image data from the SD card in its memory. After that, a convolutional neural network (CNN) is used to process the image. The FPGA is used to process the entire image. Finally, FPGA communicates with the VGA interface via a VGA connection in order to display the processed image on the monitor. In this work we try to implement CNN architecture in FPGA hardware. Our work is implemented in the Altera DE 1 FPGA board of the EP2C20F484C7N device, which belongs to the Cyclone II family.

Keywords—Field Programmable Gate Array, FPGA, Deep learning, field hockey, Verilog, Convolutional Neural Network, CNN, object tracking, sports

I. INTRODUCTION

In the past few years, deep learning has become one of the most popular research directions in the computer field. In particular, Convolutional Neural Networks (CNN) have performed well in application areas such as image classification, recognition, detection, and segmentation, and have attracted more and more attention [13]. The success of

deep learning (DL) has fast paced the evolution of current technology at unprecedented rate. In particular, deep convolutional neural networks (CNNs) has gained a lot of attention due to their extraordinary performance in a wide range of computer vision applications. The implementation complexity of deep convolutional neural network, on the other hand, sets a great challenge to the smooth utilization of deep learning in many resources constrained hardware [14]. A dedicated processor is needed for implementing image processing algorithms. It is possible with VLSI technology. One of the VLSI hardware is FPGA [18]. Due to the high-performance, high-energy efficiency, reconfigurability, and low latency of FPGA, deploying CNN on FPGAs has more advantages than other hardware platforms (such as CPU, GPU, ASIC, ARM) [13]. FPGA is a type of semiconductor logic chip which can be programmed to become any kind of digital circuit. It's an array of logic gates. [18]. FPGA has become the preferred hardware platform for accelerating the forward inference phase of deep convolution neural network (CNN), but deploying CNN on FPGA also encountered many challenges. The implementation of CNN needs to perform complex matrix multiplication and memory access operations. The bottleneck lies in how to improve the calculation speed and memory bandwidth [13]. Convolutional neural network (CNN) is a fast forward neural network. It is used to analyze the visual image by processing data. It will extract the features of an image and classifies the image. CNN will have input layer, hidden layer and output layer. Multiple hidden layers will be present in convolution neural network. First layer of hidden layer is convolution layer. Next layer is ReLu and last layer is pooling. In output layer classifier is used to classify the features [18]. Following properties of many modern high-performing CNN architectures make their hardware implementation feasible they are,

1.high regularity: all commonly used layers have similar structure(Conv3x3, Conv1x1, MaxPooling, FullyConnected, GlobalAvgPooling),



2. small size of convolutional filters: 3×3 ,
3. ReLU activation function (comparison of the value with zero): easier to compute compared to previously used Sigmoid and Tanh functions.

Our work focuses on the employment of FPGA and deep learning techniques to improve object tracking in the fast-paced game of hockey. We intend to create an effective and precise system for tracking things, such as players and the ball, in real-time during hockey games by utilizing the strength of deep learning and the reconfigurability of FPGAs. Object tracking plays a critical role in comprehending the dynamics of hockey. The complexity of a fast-moving sport like hockey frequently makes traditional tracking technologies ineffective. In light of this, we suggest using deep learning methods, particularly Convolutional Neural Networks (CNNs), to address the issues with real-time object tracking in this situation. Field Programmable Gate Arrays (FPGAs) are a great option for object tracking in sports since they also provide considerable advantages for real-time processing activities. FPGAs are excellent at parallel processing, and they may be rearranged to improve speed and adjust to shifting tracking requirements. Our research intends to design a system that can quickly and effectively handle massive volumes of visual data and execute object tracking tasks using FPGA technology. Main objective of our work is to implement the CNN architecture in FPGA.

The rest of this article is organized as follows. Section II presents the related works. Section III explains the methodology of our work. Section IV briefly explains the detail analysis of CNN and its layers we used. Section V shows the results and its discussions. Section VI summarizes the conclusion from this work and gives the further research directions.

II. RELATED WORK

Literature survey for our work is based on CNN implementation in FPGA hardware for sports. In this paper [1] Object detection, player identification with 93-96% accuracy is done using machine learning algorithms and calculated parameters like yellow/red cards, position, poses, recall. In this study [4] Tracking all players and analyze each player's performance is done in soccer game using algorithms like YOLO, DeepSort and calculated parameters like passes, steals, player identification with accuracy of 84.81%. This study proposes movement activity and alternative point of view in football using CNN algorithm, it is implemented in Zynq FPGA board and calculated parameters like specificity, error, accuracy, sensitivity with accuracy of 99%. This study [6] proposes Appearance of players ball passing and blocking in basketball using CNN algorithm, it is implemented in Virtex FPGA board and calculated parameters like sensitivity, average, shoot, pass, catch with accuracy of 35%. This study [8] proposes Violence detection violent action recognition in hockey using Deep CNN with transfer learning algorithm, it calculated parameters like size of image, error

with accuracy of 99%. This study [13] proposes FPGA implementation of CNN to recognize hand written digits with accuracy of 97.57% using CNN algorithm, it is implemented in Intel cyclone10 FPGA board and calculated parameters like accuracy, frequency, time. This study proposes Implementation of CNN architecture in FPGA using CNN algorithm, it is implemented in Altera (Cyclone IV) FPGA board and it calculated parameters like speed. This study [14] proposes Neural network architecture implementation in FPGA using CNN algorithm, it is implemented in Zynq FPGA board and it calculated parameters like efficiency, accuracy, size reduction. This study [15] proposes Implementation of CNN architecture in FPGA to classify hyperspectral images using CNN algorithm, it is implemented in Altera (Cyclone IV) FPGA board and it calculated parameters like accuracy and speed. This study [16] proposes Implementation of CNN architecture in FPGA to solve power consumption problem using CNN algorithm, it is implemented in Zynq FPGA board and it calculated parameters like memory and power. This study [17] proposes Implementation of CNN architecture in FPGA for power efficiency to enable image recognition using CNN algorithm, it is implemented in Altera DE-5 FPGA board and it calculated parameters like power, time, memory performances. [5] proposed a system which can achieve a higher throughput and higher energy efficiency. The S-BOX is designed by using Advanced Encryption Standard (AES). The AES is a symmetric key standard for encryption and decryption of blocks of data. In encryption, the AES accepts a plaintext input, which is limited to 128 bits, and a key that can be specified to be 128 bits to generate the Cipher text. In decryption, the cipher text is converted to original one. By using this AES technique the original text is highly secured and the information is not broken by the intruder. From that, the design of S-BOX is used to protect the message and also achieve a high throughput, high energy efficiency and occupy less area. [7] proposed a system which contributes the complex parallelism mechanism to protect the information by using Advanced Encryption Standard (AES) Technique. AES is an encryption algorithm which uses 128 bit as a data and generates a secured data. In Encryption, when cipher key is inserted, the plain text is converted into cipher text by using complex parallelism. Similarly, in decryption, the cipher text is converted into original one by removing a cipher key. The complex parallelism technique involves the process of Substitution Byte, Shift Row, Mix Column and Add Round Key. The above four techniques are used to involve the process of shuffling the message. The complex parallelism is highly secured and the information is not broken by any other intruder.

The prior research demonstrates that utilizing CNN provides good accuracy and also demonstrates that FPGA implementation for hockey games doesn't occur. In order to recognize and track objects during a hockey game, the CNN

HARDWARE SPECIFICATIONS	
Board	Altera DE1
Family	Cyclone II
Device	EP2C20F256C8N
Memory	SDRAM (8 Mbytes)
Logic elements	68,416
I/O Pins	622
SD-card	Capacity-32GB
SD-card I/O pins	8 pins
VGA	Monitor screen resolution: 1366x768-60Hz

algorithm is implemented in the Altera DE1 FPGA board in our work.

III. METHODOLOGY

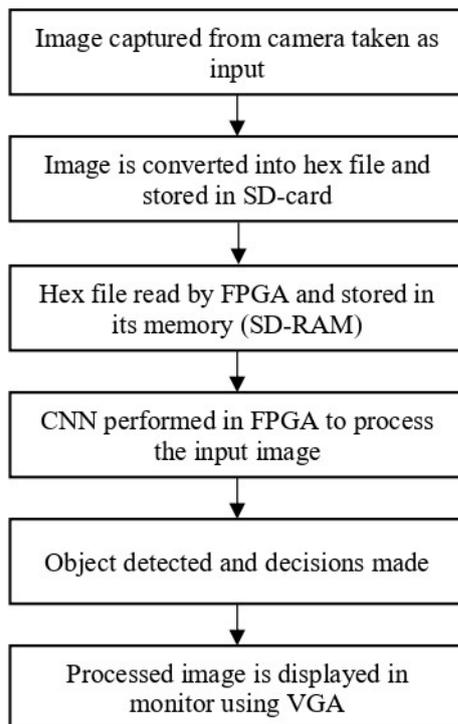


Fig.1. Flow chart of methodology

This work uses an FPGA to implement one of the deep learning algorithms, CNN, for hockey object tracking to improve player's performance for that ball predictions are needed at a time of practice to find player strategies, team strategies. The implementation of CNN architecture on an

FPGA is the primary goal of our effort. The dataset was initially designed for hockey. The next step is to use CNN to extract the objects, such as hockey sticks and balls. The next step is to track objects in order to make decisions. The image is taken by the camera, transformed to a hex format, and then saved on the SD card. To connect with the camera and FPGA, an SD-card interface has been developed. The FPGA reads the hex file from the SD card and stores it in its memory. After that, CNN is used to process the input image and make decisions. Then, using a VGA interface and a VGA cable, the processed image and its conclusions are shown on the monitor. To communicate with the FPGA and the monitor, a VGA interface is developed. Fig.1. provides the overall flow chart for the approach used in this study. This work is the continuation of this paper [18]. In this paper [18] all the interfacing and preprocessing works are done. So, the main objective of this paper is to implement CNN architecture in FPGA using Verilog HDL. So first we created module hierarchy in Quartus software for CNN and its layers and according to these modules we done implementations.

Hardware Specification:

TABLE I HARDWARE SPECIFICATION

The hardware specifications of the device we are using are mentioned in TABLE I.

Altera FPGA board:

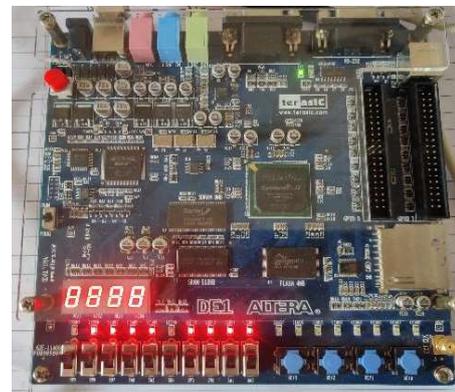


Fig.2 Altera DE1 FPGA board

In our work, for hardware implementation, we are using Altera DE1 FPGA board which is shown in Fig.2

IV. DETAILED ANALYSIS

Our work flows as follows: preprocessing of image data; interacting the data on the SD card with the FPGA; processing the image using the CNN algorithm to detect and track the objects; and finally displaying the processed data from the FPGA to monitors using the VGA interface. Since our work is a continuation of our previous work [18], we finished all the preprocessing and interfacing in that paper.

So, the detailed analysis of interfaces is mentioned in the cited paper [1]. In this paper, we only focused on CNN, so in this section, a detailed analysis of CNN will be discussed.

CNN Hierarchy:

In Quartus II software, for writing a Verilog HDL code modules and its submodules will be created. It is generally called hierarchy of the module. Based on this hierarchy, software implementation for the module and its layers are generated.

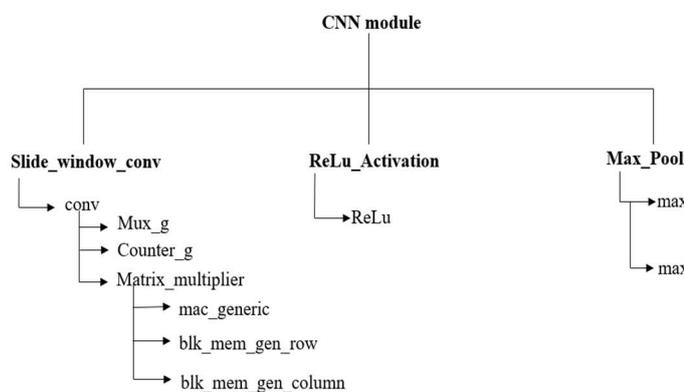


Fig.3 CNN module hierarchy in Verilog HDL

CNN module hierarchy in Verilog HDL is explained in Fig.3

Pin Configuration of CNN and its layers:

Pin configuration for CNN and its layers is explained here.

CNN module:

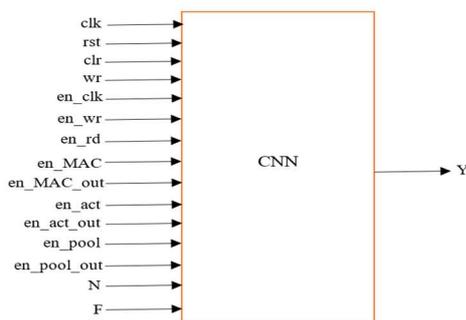


Fig.4 CNN module Pin diagram

A fast-forward neural network is CNN. It will classify the image after extracting its features. The pin diagram for CNN is mentioned in Fig.4

Slide_window_conv:

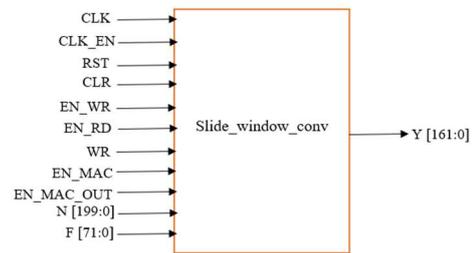


Fig.5 Sliding window layer Pin diagram

The sliding window layer in CNNs slides a filter over input data, extracting features through multiplications and summations. These features form spatially-aware maps, helping the network learn intricate patterns. The pin diagram for sliding window layer is mentioned in Fig.5

ReLU_Activation:

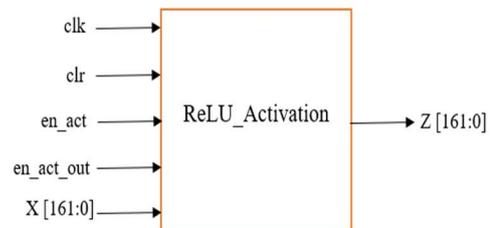


Fig.6 ReLU activation layer Pin diagram

The ReLU activation layer introduces non-linearity by setting negative values to zero and keeping positive values unchanged, enhancing network performance and enabling learning of complex patterns. The pin diagram for ReLU activation layer is mentioned in Fig.6

Max_Pooling:

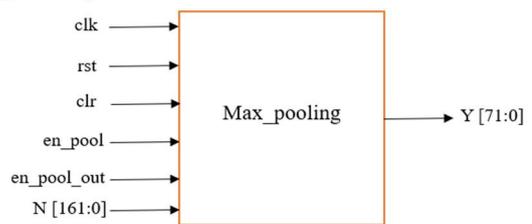


Fig.7 Max_Pooling layer Pin diagram

The max pooling layer reduces the spatial dimensions of feature maps by selecting the maximum value within each pooling window, helping to extract the most salient features while preserving important spatial information in the network. The pin diagram for Max Pooling layer is mentioned in Fig.7

Pin Configuration of Sub modules:

The pin diagram for each layer's sub module is explained here.

Slide_window_conv sub modules:

The sub modules of Slide_window_conv is conv, mux, counter, matrix multiplier layer.

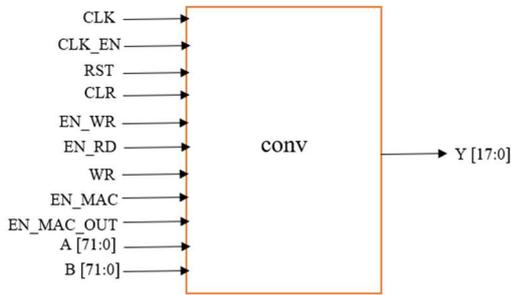


Fig.8 Conv Pin diagram

The pin diagram for conv sub module is mentioned in Fig.8

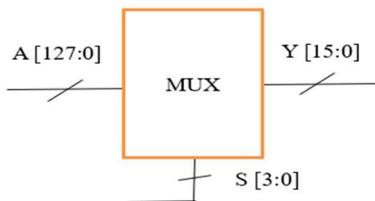


Fig.9 Mux Pin diagram

The pin diagram for mux sub module is mentioned in Fig.9

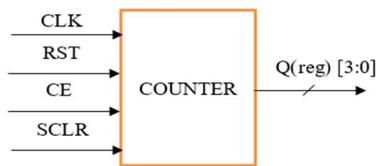


Fig.10 counter Pin diagram

The pin diagram for counter sub module is mentioned in Fig.10

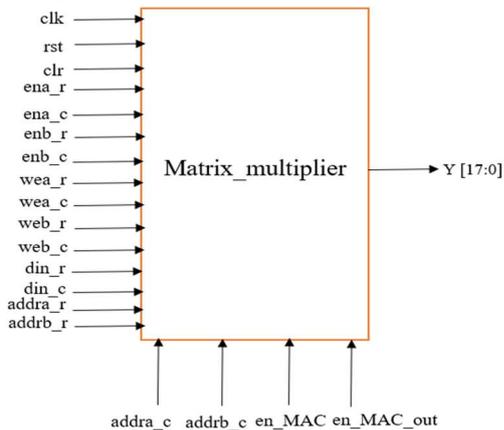


Fig.11 matrix multiplier Pin diagram

The pin diagram for matrix multiplier sub module is mentioned in Fig.11

ReLU_Activation sub modules:

The sub module of ReLU_Activation is ReLU layer.

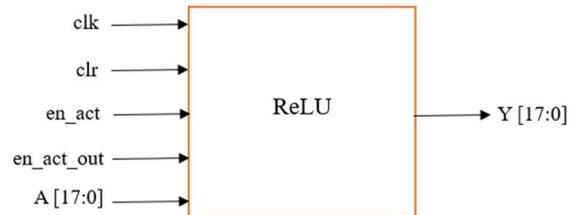


Fig.12 ReLU Pin diagram

The pin diagram for ReLU sub module is mentioned in Fig.12

Max_Pooling sub modules:

The sub modules of Max_Pooling is max data and max value.

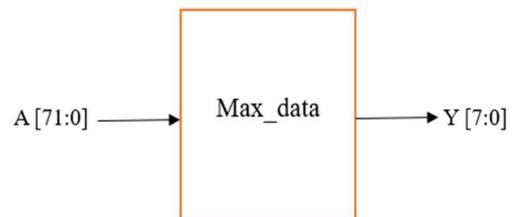


Fig.13 max data Pin diagram

The pin diagram for max data sub module is mentioned in Fig.13

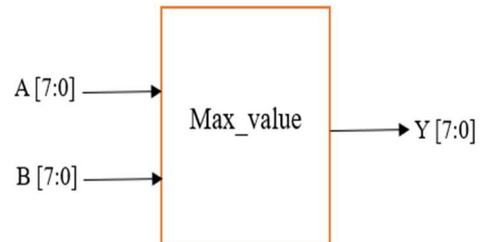


Fig.14 max value Pin diagram

The pin diagram for max value sub module is mentioned in Fig.14

V. RESULTS AND DISCUSSION

The simulation results of each layer and its sub-module in the CNN architecture hierarchy are explained in this block. The output waveform is generated according to the given inputs in the software simulation. The verilog codes are compiled using Quartus II software, and their waveform is verified using ModelSim software.

Mux_g:

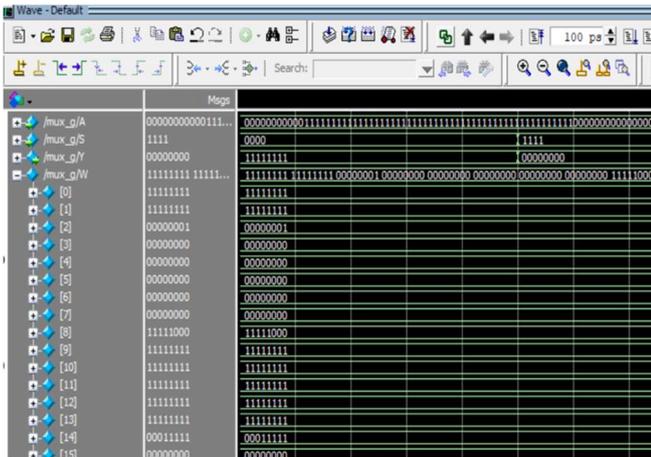


Fig.15 Mux_g simulation result

Mux_g simulation result is given in Fig.15; the output is determined by the select signals. When the select signal is 0, the output is the value of D0 (11111111). When the select signal is 15, the output is the value of D15 (00000000).

Counter_g:



Fig.16 Counter_g simulation result

Counter_g simulation result is given in Fig.16; the output generated based on counter enable signal, CE. If CE=1, it is incremented and count the value and store in Q. And if CE=0, it will not count the value and remains unchanged.

Mac_generic:

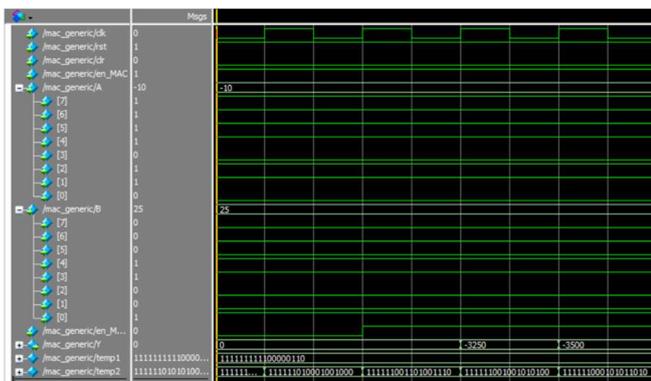


Fig.17 Mac_generic simulation result

Mac_generic simulation result is given in Fig.17; value of A and B is multiplied and accumulated and stored in temp 1 and 2. Output, Y is generated based on en_MAC and en_MAC_out values.

Blk_mem_gen_ROW:

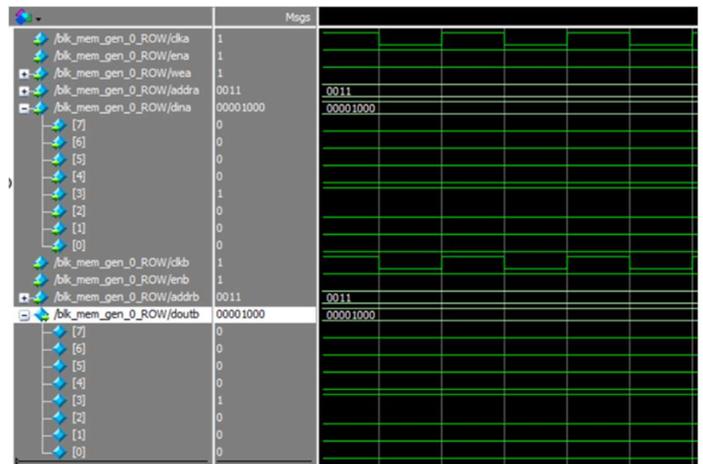


Fig.18 Blk_mem_gen_ROW simulation result

Blk_mem_gen_ROW simulation result is given in Fig.18; when all write enable and read enable signal high, read and write operation happens. The given din got by doutb which shows row block generated.

Blk_mem_gen_COLUMN:

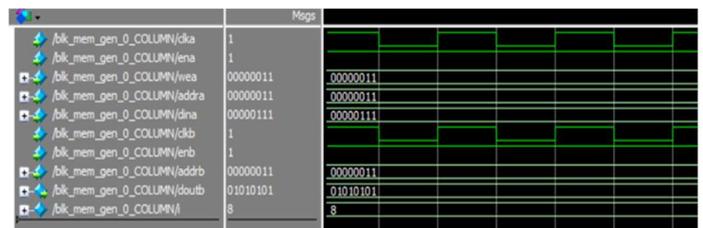


Fig.19 Blk_mem_gen_COLUMN simulation result

Blk_mem_gen_COLUMN simulation result is given in Fig.19; when all write enable and read enable signal high, read and write operation happens. Based on memory generation we got doutb which shows column block generated.

ReLU_Activation:

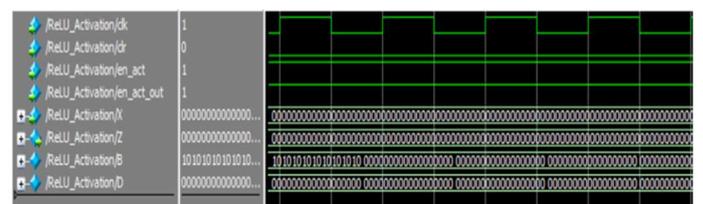


Fig.20 ReLU activation simulation result

ReLU activation simulation result is given in Fig.20; the input is given in X and it stores the value in B and its instance ReLU module performs cases in ReLU and store in D and that value is stored in Z given as output.

ReLU:

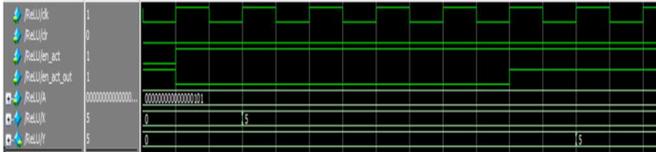


Fig.21 ReLU simulation result

ReLU activation simulation result is given in Fig.21; input given as 5. So, output is generated as 5. If input is negative ReLU layer will replace it with 0.

Max_value:

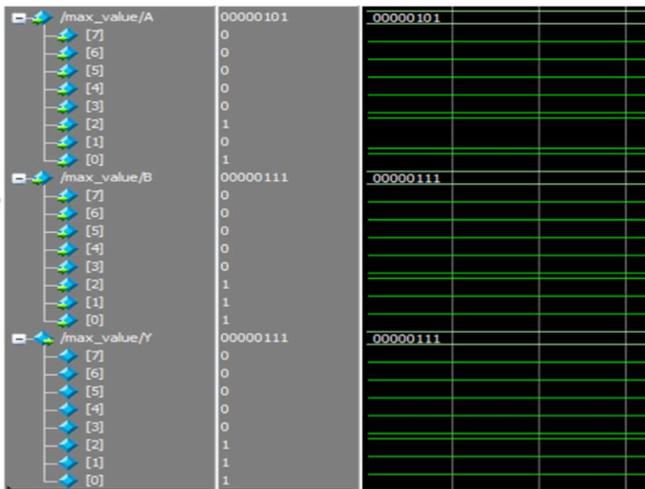


Fig.22 Max_value simulation result

Max_value simulation result is given in Fig.22; it compares both value and give output as maximum value.

Max_data:

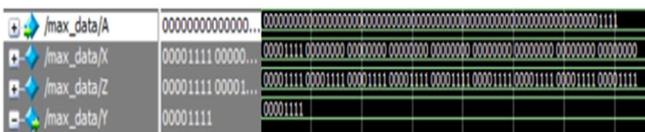


Fig.23 Max_data simulation result

Max_data simulation result is given in Fig.23; it performs element-wise comparison and generates output.

Max_Pooling:

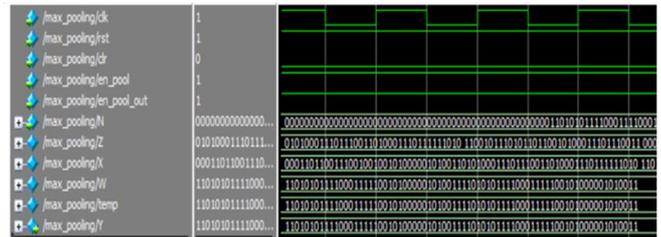


Fig.24 Max_Pooling simulation result

Max_Pooling simulation result is given in Fig.24; based on padding, stride, pooling value it performs Max_pooling-selects maximum value of feature maps.

The results for the software implementation of all layers and their submodules are generated except slide window, conv, and matrix multiplier. We are still working on these three submodules.

VI. CONCLUSION AND FUTURE WORK

This work presents Deep learning-based object tracking in field hockey using FPGA. This work provides an overview of FPGA with various image processing techniques, the development of some new technologies, real-time application solutions, and various implementations of earlier literature. This paper also explains shortcomings, upcoming projects, and literary implications from earlier works. Hierarchy of CNN module is generated. Pin configurations of each layer of CNN and its submodule is analyzed and made. Software simulation of all layers and sub module of CNN module expect three sub modules is generated using Quartus II and ModelSim software using Verilog HDL. For future work, our goal is to Implement the CNN architecture using an Altera DE1 hardware board Next is to process the image using CNN with the help of datasets to detect ball in an image for ball predictions. Then object tracking is performed in videos for making decisions automatically.

ACKNOWLEDGMENT

This publication reflects purely the author's view and use of the information contained in the paper is the pure responsibility of the user.

REFERENCES

[1] Naik B.T.; Hashmi, M.F.; Bokde, N.D. A Comprehensive Review of Computer Vision in Sports:Open Issues, Future Trends and Research Directions. Appl. Sci. 2022, 12, 4429.
[2] R. Neris, A. Rodríguez, R. Guerra, S. López and R. Sarmiento, "FPGA- Based Implementation of a CNN Architecture for the On-Board Processing of Very High-Resolution Remote Sensing Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 3740-3750, 2022, doi: 10.1109/JSTARS.2022.3169330.

- [3] M. -S. Kang, R. -H. Park and H. -M. Park, "Efficient Spatio-Temporal Modeling Methods for Real-Time Violence Recognition," in *IEEE Access*, vol. 9, pp. 76270-76285, 2021, doi: 10.1109/ACCESS.2021.3083273.
- [4] R. Theagarajan and B. Bhanu, "An Automated System for Generating Tactical Performance Statistics for Individual Soccer Players From Videos," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 632-646, Feb. 2021, doi: 10.1109/TCSVT.2020.2982580.
- [5] Christo Ananth, H. Anusuya Baby, "S-Box using AES Technique", *International Journal of Engineering Research & Technology (IJERT)*, Vol. 3 Issue 3, March – 2014, pp 285-290
- [6] Gun Junjun, "Basketball action recognition based on FPGA and particle image", *Microprocessors and Microsystems*, Volume 80, 2021, 103334, ISSN 0141-9331
- [7] Christo Ananth, H. Anusuya Baby, "High Efficient Complex Parallelism for Cryptography", *IOSR Journal of Computer Engineering (IOSR-JCE)*, Volume 16, Issue 2, Ver. III (Mar-Apr. 2014), PP 01-07.
- [8] A. Mumtaz, A. B. Sargano and Z. Habib, "Violence Detection in Surveillance Videos with Deep Network Using Transfer Learning," 2018 2nd European Conference on Electrical Engineering and Computer Science (EECS), Bern, Switzerland, 2018, pp. 558-563, doi: 10.1109/EECS.2018.00109.
- [9] Wang, Shang and Zhang, Chen and Shu, Yuanchao and Liu, Yunxin, "Live Video Analytics with FPGA-Based Smart Cameras", 2019, Association for Computing Machinery.
- [10] S. V. Mora and W. J. Knottenbelt, "Deep Learning for Domain-Specific Action Recognition in Tennis," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 170-178, doi: 10.1109/CVPRW.2017.27.
- [11] R. Zeng, R. Lakemond, S. Denman, S. Sridharan, C. Fookes and S. Morgan, "Vertical Axis Detection for Sport Video Analytics," 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, Australia, 2016, pp. 1-7, doi: 10.1109/DICTA.2016.7797093.
- [12] Punith Kumar M B and P. S. Puttaswamy, "The extraction of events and replay in cricket video," 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2015, pp. 54-58, doi: 10.1109/ERECT.2015.7498987.
- [13] R. Xiao, J. Shi and C. Zhang, "FPGA Implementation of CNN for Handwritten Digit Recognition," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2020, pp. 1128-1133, doi: 10.1109/ITNEC48623.2020.9085002.
- [14] M. Hailesellasiye, S. R. Hasan, F. Khalid, F. A. Wad and M. Shafique, "FPGA-Based Convolutional Neural Network Architecture with Reduced Parameter Requirements," 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 2018, pp. 1-5, doi: 10.1109/ISCAS.2018.8351283.
- [15] X. Chen, J. Ji, S. Mei, Y. Zhang, M. Han and Q. Du, "FPGA Based Implementation of Convolutional Neural Network for Hyperspectral Classification," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 2018, pp. 2451-2454, doi: 10.1109/IGARSS.2018.8517973.
- [16] S. Hareth, H. Mostafa and K. A. Shehata, "Low power CNN hardware FPGA implementation," 2019 31st International Conference on Microelectronics (ICM), Cairo, Egypt, 2019, pp. 162-165, doi: 10.1109/ICM48031.2019.9021904.
- [17] M. Bettoni, G. Urgese, Y. Kobayashi, E. Macii and A. Acquaviva, "A Convolutional Neural Network Fully Implemented on FPGA for Embedded Platforms," 2017 New Generation of CAS (NGCAS), Genova, Italy, 2017, pp. 49-52, doi: 10.1109/NGCAS.2017.16.
- [18] M. Praveena, V. V. Thyagarajan and V. R. Venkatasubramani, "Video Analytics with FPGA based smart cameras for Object Recognition in the game of field hockey," 2023 International Conference on Networking and Communications (ICNWC), Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICNWC57852.2023.10127549.