Human Face Detection Using Compact CNN Accelerator IP

Reference design supported by Neural Network Compiler, CNN IP, or Compact CNN IP 1.1. Not recommended for new designs.

Reference Design

FPGA-RD-02034-1.2

May 2019

Disclaimers

Lattice makes no warranty, representation, or guarantee regarding the accuracy of information contained in this document or the suitability of its products for any particular purpose. All information herein is provided AS IS and with all faults, and all risk associated with such information is entirely with Buyer. Buyer shall not rely on any data and performance specifications or parameters provided herein. Products sold by Lattice have been subject to limited testing and it is the Buyer's responsibility to independently determine the suitability of any products and to test and verify the same. No Lattice products should be used in conjunction with mission- or safety-critical or any other application in which the failure of Lattice's product could create a situation where personal injury, death, severe property or environmental damage may occur. The information provided in this document is proprietary to Lattice Semiconductor, and Lattice reserves the right to make any changes to the information in this document or to any products at any time without notice.

Contents

Figures

1. Introduction

The Lattice Human Face Detection using Compact CNN Accelerator Reference Design document describes how to implement a Face Detection design on our iCE40™ UltraPlus FPGA. This design utilizes the Lattice Radiant Compact CNN Accelerator IP core which is optimized for Binary Neural Network implementations.

2. Related Documentation

In addition to using this guide to help you get started developing Compact CNN solutions on your device, you can refer to other applicable documents that may contain more detailed information that is beyond the scope of this guide.

The following documents can be obtained in the Lattice website:

- Lattice SensAI Neural Network Compiler Software User Guide (FPGA-UG-02052) – This document explains how to create the firmware file which contains the command sequence, as well as weights that go into the Compact CNN Accelerator IP Core.
- Compact CNN Accelerator IP Core User Guide (FPGA-IPUG-02038) – This document goes into detail on everything you need to know about the Compact CNN IP Core contained in this design.
- Lattice Radiant Software User Guide - This reference design is designed in Lattice Radiant software, and this user guide explains everything you need to know about Lattice Radiant.

3. Software Requirements

The following software requirements are required in order to design and use the Face Detection Reference Design:

- Radiant Software 1.0 or greater – This is used to build the reference design.
- Radiant Programmer tool – This is used to program the bitstream to the External SPI Flash on the board.
- Compact CNN Accelerator IP Core – This is used to create the Compact CNN Accelerator Engine.
- Lattice SensAI Neural Network Compiler Software – This is used to generate your firmware file from your TensorFlow/Caffe outputs.

4. Hardware Requirements

The reference design has been targeted to support the Mobile Development Platform board, but can be modified for use in other HW as needed.

5. Directory Structure

Figure 5.1 shows the directory structure when unzipping the Implementing Human Face Detection Using Compact CNN Accelerator IP Source Code. The figure explains what files are contained in each folder.

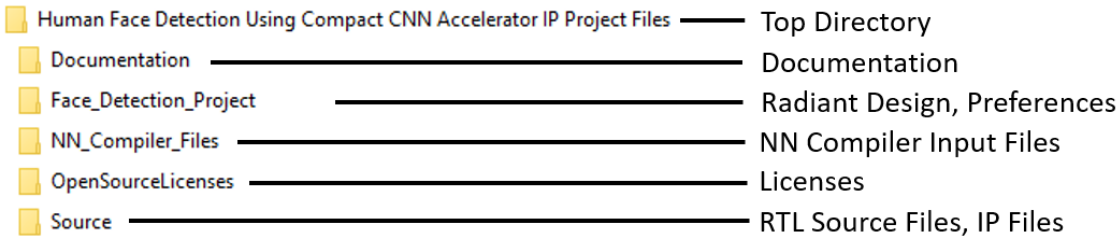

Figure 5.1. Directory Structure of Implementing Face Detection

6. Human Face Detection Reference Design Overview

The purpose of this reference design is to show an implementation of a machine learning design with our Compact CNN Accelerator Soft IP Engine. The Face Detection reference design will be first described with a block diagram of the design, followed by a description for each top block module.

6.1. Block Diagram

The block diagram is separated into white and gray. White blocks are the main top level module implemented on the iCE40 UltraPlus, while the gray blocks are external parts. The dotted line represents the boundaries of the iCE40 UltraPlus FPGA. There are two main inputs that go into the Compact CNN Accelerator Engine, which then outputs 2 values that get evaluated in order to determine whether there is a face or not.

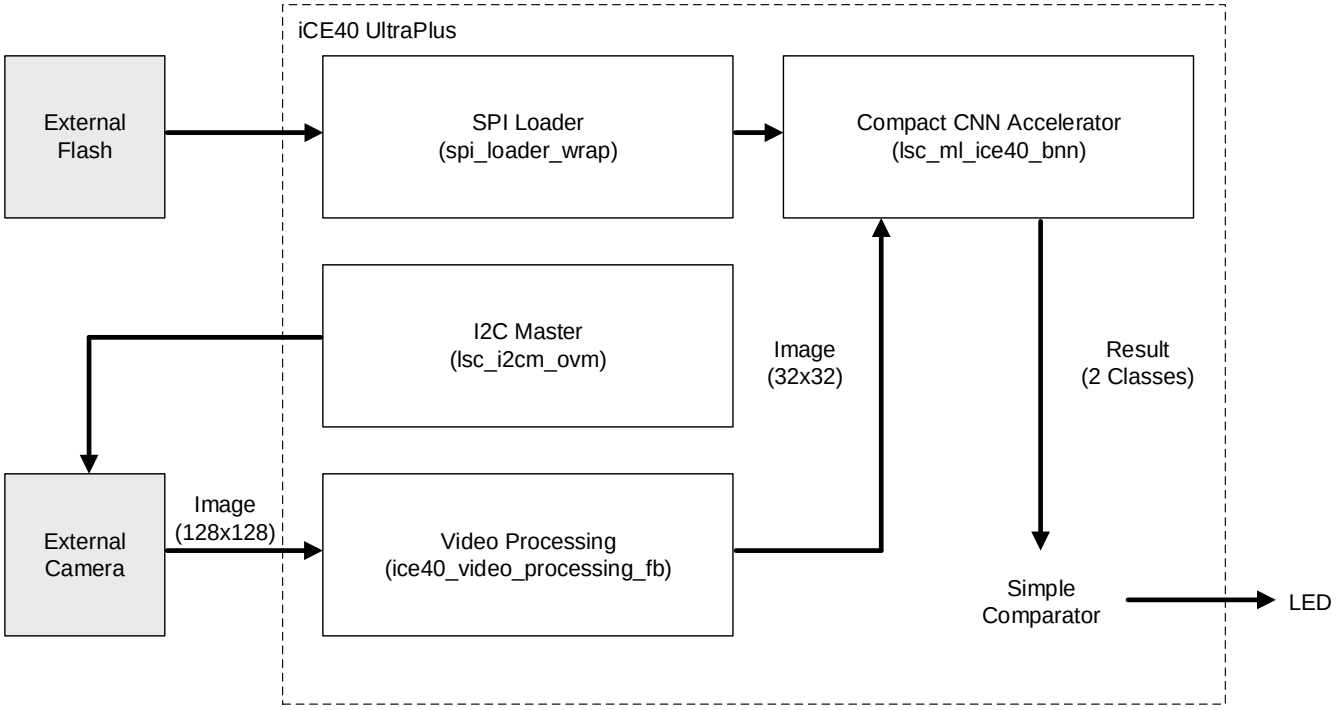

Figure 6.1. Human Face Detection Block Diagram

6.2. Top Level Blocks

This section discusses the details of each of the main top level modules that makes up the Human Face Detection Reference design. Each top module includes which Verilog file(.V) or IP catalog (.IPX) file it contains. These files can all be found in the source directory of the reference design package.

6.2.1. Top Level – lsc_ml_ice40_face_det_top.v

This is the top level module which contains all the blocks and connections found in Figure 6.1.

6.2.2. Compact CNN Accelerator Engine – lsc_ml_ice40_bnn.ipx

This block contains the Compact CNN Accelerator engine. In this design, the engine is configured to Machine Learning Type as BNN, Memory Type as Single_SPRAM and BNN Blob Type as +1/0. To understand what these mean, please refer to the Compact CNN Accelerator IP Core User Guide (FPGA-IPUG-02038). This ipx file takes the Firmware file from the External SPI Flash and the image from the External Camera to output 2 values, Face or No Face. This firmware file is generated by the Lattice Neural Network complier tool. The input to the tool are the .caffemodel and .proto file. These files are located in the NN_Compiler_Files directory for you to generate your own firmware file.

6.2.3. SPI Loader – spi_loader_wrap.v (top block) and spi_loader_spram.v

This module reads the external flash for the firmware file. In this reference design, the starting address is specified to be 20'h20000. The firmware file would have to also be programmed in the SPI Flash to start at the same address. There are multiple versions of SPI Loader provided and this Face Detection module uses spi_loader_spram.v which contains 2 SPRAM blocks. For your convenience, there is a spi_loader_single_spram.v and spi_loader_ebram.v included in the Source/Common directory, if needed when implementing your own design. Switching between different types of SPI files is dependent on the amount of memory storage on the external SPI Flash. If your design requires less than 8 KB of memory, you can use EBR setting in the Compact CNN Accelerator along with using spi_loader_ebram.v. This implementation will use 0 SPRAM blocks, but 16 EBR. If you do not want to use EBR for storage, there is an option to use SPRAM instead. Single SPRAM configuration along with spi_loader_single_spram.v can store up to 32 KB of data, while Dual SPRAM with spi_loader_spram.v can store up to 64 KB of data.

6.2.4. I²C Master – lsc_i2c_ovm.v (top block) and lsc_i2cm.v

The I²C master block controls the external camera. When the board powers up, the camera needs to be initialized first. The I²C Master module has a RAM IP block that contains a memory file named ram256x16_ovm.mem, which contains the initialization data in the EBR which is then transferred to configure the Camera settings. The camera is then configured to produce 8-bit 128 x 128 images which is then inputted into the FPGA for processing. After successfully initializing the external camera, the clock generator module (refer to Clock Generator section) turns off this module to save power.

6.2.5. Video Processing – ice40_video_process.v

Human Face Detection is fairly simple and does not need a very high resolution image. This module takes in 8-bit 128 x 128 image data and scales it down 4 times to a 32 x 32 image to help our Compact CNN Accelerator soft IP increase performance without sacrificing much accuracy. This 32 x 32 image is then fed into the Compact CNN Accelerator soft IP. The video processing module not only gives the Accelerator image data, but also creates the addresses and the write enable signal which are also fed into the Accelerator Engine. In Figure 6.2, it shows how the data and addresses are transferred into the Compact CNN Accelerator. The data is fed in order of Blue, Green, and then Red data. Each color will take 1024 addresses (32 x 32), which is a total of 3072 addresses.

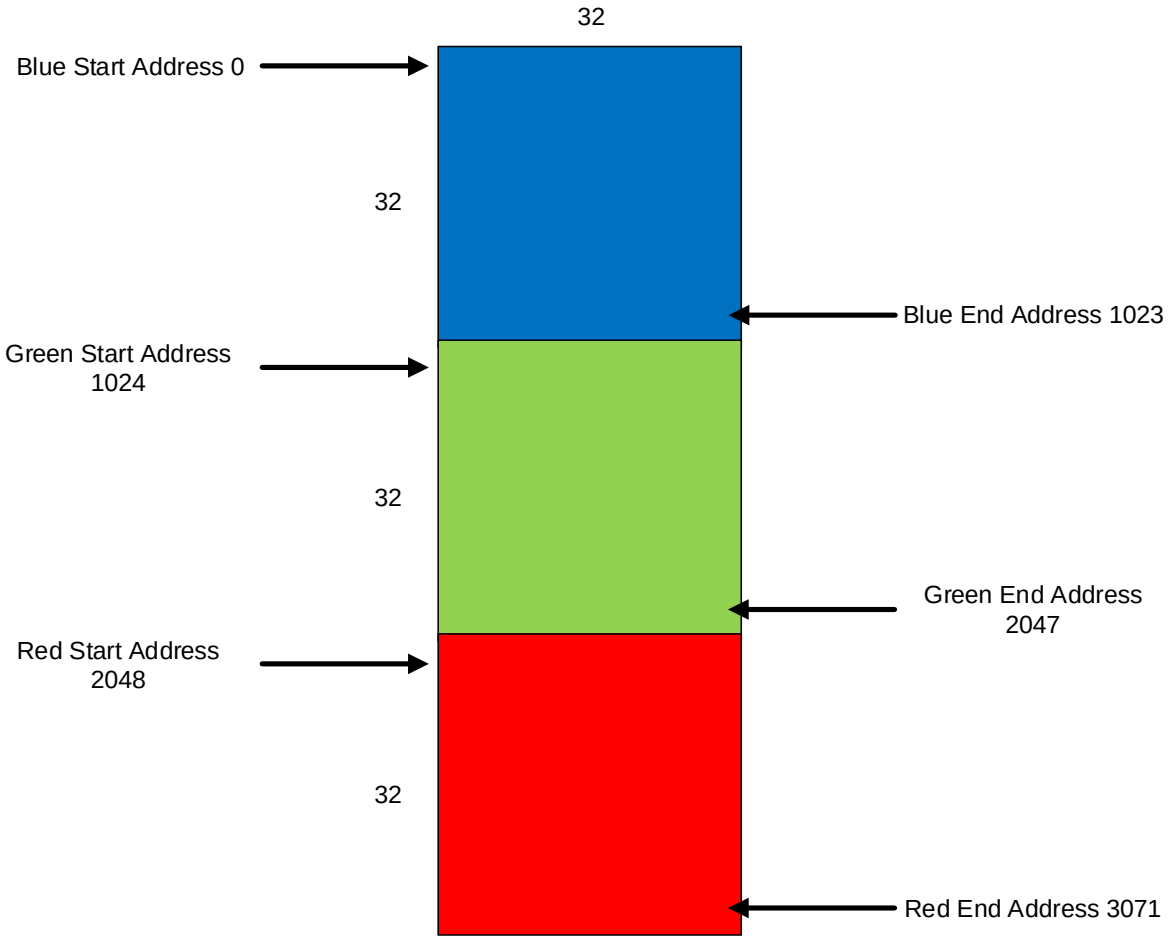

Figure 6.2. How Data/Address is fed into Compact CNN Accelerator for Human Face Detection

6.2.6. Clock Generator – ice40_facedet_clkgen.v

This module is not included in Figure 6.1, but is still relevant. This module generates and controls all the clocks in the design. Whenever a module is not used, this module will stop transferring the clock signal to that module in order to save power. For example, after the i2cm block initializes the external camera, this module will cut off the clock to that module since there is no use for that module after initialization.

6.2.7. Comparator – Implemented in top level logic

The comparator is a simple algorithm that takes the difference of computed value of face and no face. That value is then compared with a threshold value. If the difference is greater than threshold value, the LED stays or turns on. Otherwise, the LED stays off or turns off. This threshold value can be changed in the RTL code as signal w_diff_th.

7. Generating the Firmware File

To generate the Human Presence Firmware file:

1. Using the files located in the **NN_Compiler_Files** directory, create a new project in SensAI and apply the following settings as shown in Figure 7.1.
   - **Framework – Caffe**
   - **Device – Ultra Plus**
   - **Class – BNN**
   - **Network File – facedet3_bnn.proto**
   - **Model File – facedet3_bbn.caffemodel**
   - **Image/Video/Audio Data – man2.jpg**
2. Click **Next**.

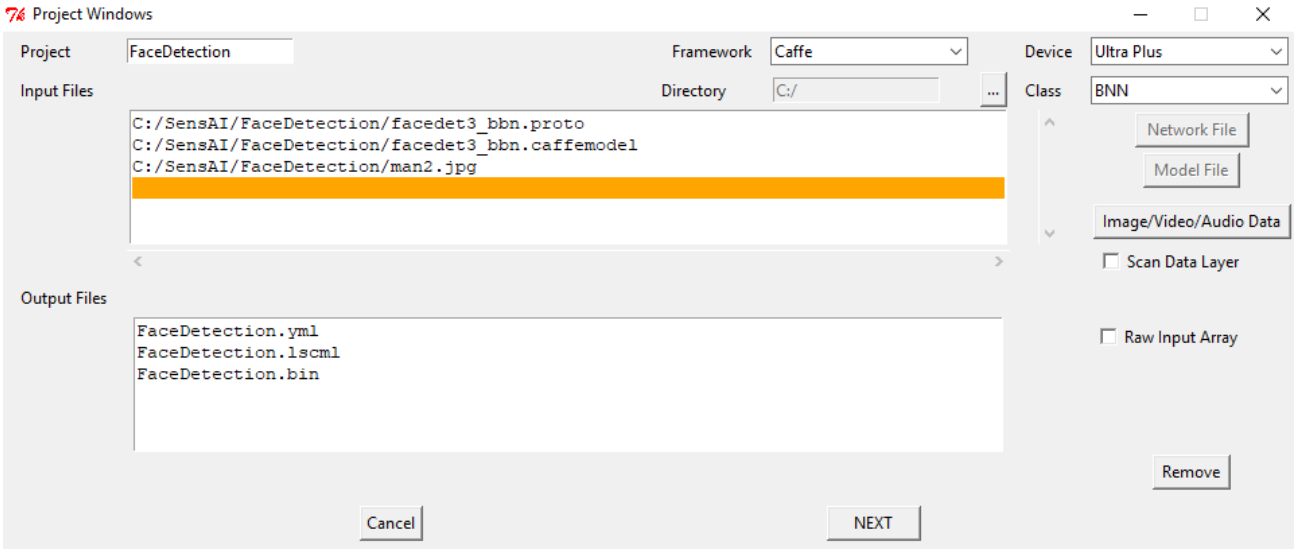

Figure 7.1 SensAI Project Settings Part 1

3. In the second section, apply the Neural Network engine settings as shown in Figure 7.2:
   - **On-Chip Memory Block Size – 32768**
   - **Mean Value – 0**
   - **Scale Value – 1.0**
   - **Quantization Mode for BNN – 0/1**
4. Click **OK**.

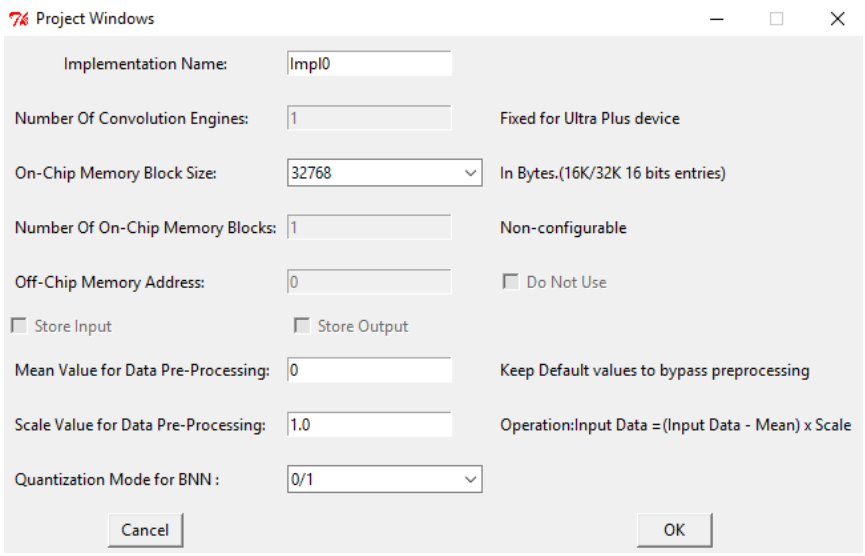

Figure 7.2 SensAI Project Settings Part 2

5. Click **Analyze & Compile** to generate the Face Detection firmware file.

Technical Support Assistance

Submit a technical support case through www.latticesemi.com/techsupport.

Revision History

Revision 1.2, May 2019

Section	Change Summary
All	• Added *Reference design supported by Neural Network Compiler, CNN IP, or Compact CNN IP 1.1. Not recommended for new designs*. on the front cover page. • Added Disclaimers section. • Updated the back cover page of the document.

Revision 1.1, September 2018

Section	Change Summary
All	• Renamed the document from Human Face Detection Using BNN Accelerator IP to Human Face Detection Using Compact CNN Accelerator IP. • Changed all instances from BNN Accelerator to Compact CNN Accelerator.
Software Requirements	Added a bullet point for Lattice SensAI Neural Network Compiler Software.
Human Face Detection Reference Design Overview	Updated Figure 6.1. Human Face Detection Block Diagram..
Generating the Firmware File	Added this section in the document.
Revision History	Updated revision history table to new template.

Revision 1.0, May 2018

Section	Change Summary
All	Initial release

www.latticesemi.com