FPGgram

Diana Valverde
Tonye Brown

Overview

- Aim
- Design
  - Neural Network Structure
  - Hardware
  - Software
- Results
- Lesson Learned

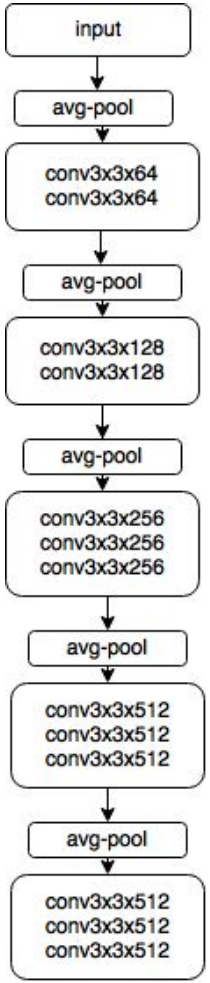
Aim

Our project focuses on using convolutional neural network for image processing. Specifically we would like to recreate an image in an artistic style. The output image is created through convolutional neural network that recognizes the content of the image and applies the style of a separate image.

Our aim is to accelerate this Very Deep Convolutional Neural Network by implementing layers of the network in hardware and allow a software program to interface between these layers.

VGG network

Our project implemented the convolution and average pool functions as well as additional units to handle the back-propagation



Top level

- 64 bit ISA
- Memory control instruction
  - High 32 bits: 1 bit (mem or ALU), 3 bits (buffer to read/write), 1 bit (reset), 26 bits DDR3 address
  - Low 32 bits: 8 bits (stride), 8 bits (rows), 16 bits (block)
- ALU instruction
  - High 32 bits: 1 bit (mem or ALU), 3 bits (buffer to read), 4 bits (output sub ID), 2 bits (input/output sub block), 2 bits (read row + whether it's a row or column - also use input/output sub block), 1 bit reverse mask), 19 empty bits
  - Low 32 bits: 16 bits (input block ID), 16 bits (output block ID)
- 128 bit data - may used shared memory to make this transfer

Mem control Unit

- Read buffer 1
  - 256x256 buffer
  - Composed of 64 4x4RAMs
    - This is to access blocks of data in one clock cycle
  - Bit for padded read
- Main buffer used
- Reads image from DDR3 based on stride, row, block inputs

Mem control unit contd.

- Read buffer 2
  - 256x128 buffer
  - Composed of 4x4 RAMs
  - Bit for padded read
- Secondary buffer - only used for Gram matrix calculations and
- Reads from DDR3

Mem control unit contd.

- Mask buffer
  - 4x4 buffer
- Reads mask from memory

Write back accumulator

- Write buffer
  - 256x256
  - Composed of 16 dual-port rams
    - To write/accumulate in one cycle
- Writes results back to DDR3 once accumulation is done

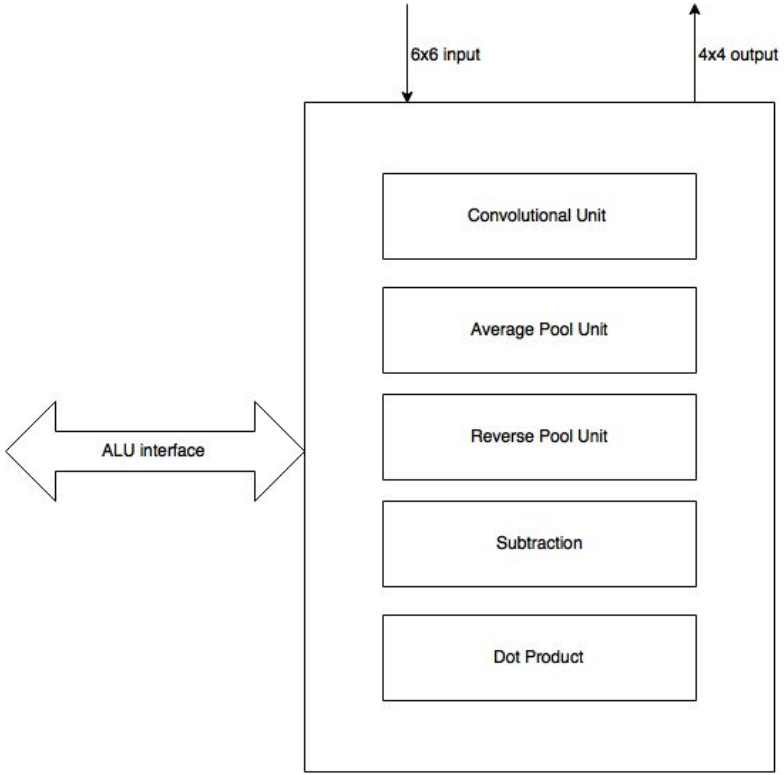
64 RAMs

- Otherwise takes 1-2 hours to compile
  - Single-cycle 4x4 block access too complicated for altera to optimize/infer RAM
- Single-cycle access made possible with striding

0	1	2	3	12	13	14	15	8	9	10	11	...	
4	5	6	7	0	1	2	3	12	13	14	15		
8	9	10	11	4	5	6	7	0	1	2	3		
12	13	14	15	8	9	10	11	4	5	6	7		
...													

ALU

- Takes in 1 64-bit input as the instruction set along with 6x6 matrix of 32 bits.
- If first bit is high instruction set goes to ALU
- 3 bits encode which buffer to read/write from.
- 4 bits for output sub id from block
- 2 bits for input/output sub-block
- 2 bits enable read and encodes whether data is in rows or columns
- 1 bit rev mask
- 16 bits encode the block ID in memory for input
- 16 bits encode the block ID in memory for output

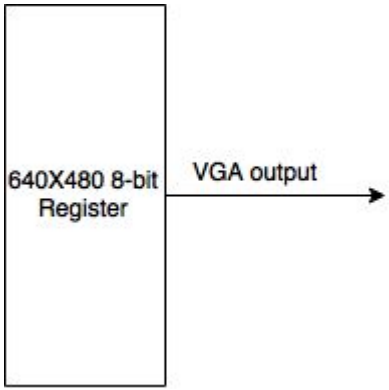


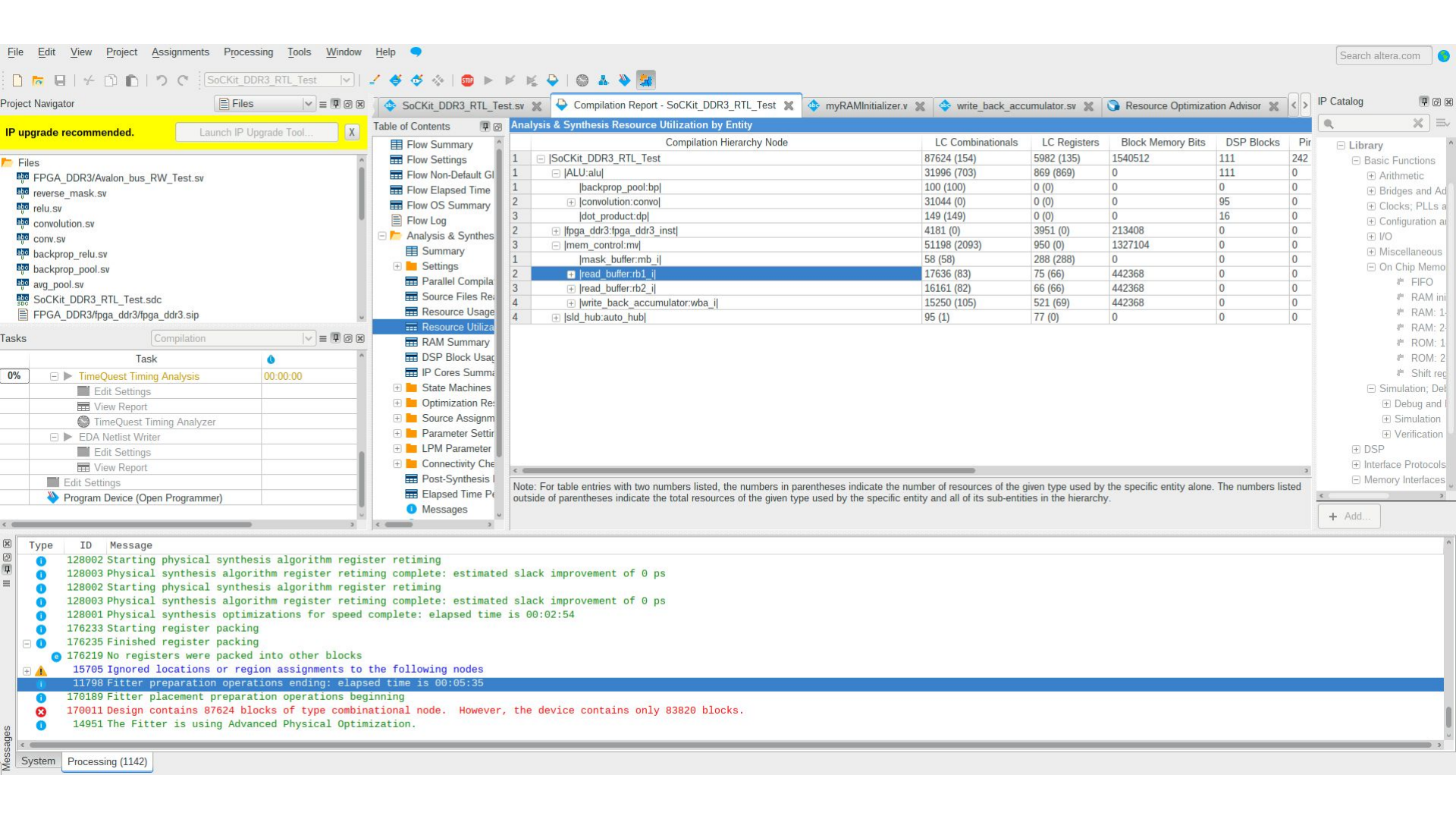
Multipliers

- 27-fixed point multiplications for ALU units
  - 1 bit signed
  - 14 bits integer
  - 13 bits fraction
- 112 multipliers on board
- 144 multipliers needed for 3x3 convolutions
- Solution: 112 hard multipliers, 22 soft multipliers

Additional Peripherals

To fully implement this project, a vga framebuffer was also implemented in the device. Pixel data is sent from CPU and stored on the framebuffer and displayed through the VGA capabilities on the FPGA




File Edit View Project Assignments Processing Tools Window Help
Search altera.com
SoCKit_DDR3_RTL_Test
Project Navigator
Files
IP upgrade recommended.
Launch IP Upgrade Tool...
Files
FPGA_DDR3/Avalon_bus_RW_Test.sv
reverse_mask.sv
relu.sv
convolution.sv
conv.sv
backprop_relu.sv
backprop_pool.sv
avg_pool.sv
SoCKit_DDR3_RTL_Test.sdc
FPGA_DDR3/fpga_ddr3/fpga_ddr3.sip
Tasks
Compilation
Task
0%
TimeQuest Timing Analysis 00:00:00
Edit Settings
View Report
TimeQuest Timing Analyzer
EDA Netlist Writer
Edit Settings
View Report
Edit Settings
Program Device (Open Programmer)
SoCKit_DDR3_RTL_Test.sv
Compilation Report - SoCKit_DDR3_RTL_Test
myRAMInitializer.v
write_back_accumulator.sv
Resource Optimization Advisor
Table of Contents
Flow Summary
Flow Settings
Flow Non-Default Gl
Flow Elapsed Time
Flow OS Summary
Flow Log
Analysis & Synthes
Summary
Settings
Parallel Compila
Source Files Re
Resource Usage
Resource Utiliza
RAM Summary
DSP Block Usag
IP Cores Summa
State Machines
Optimization Re
Source Assignm
Parameter Settir
LPM Parameter
Connectivity Che
Post-Synthesis
Elapsed Time Pe
Messages
Analysis & Synthesis Resource Utilization by Entity
Compilation Hierarchy Node | LC Combinationals | LC Registers | Block Memory Bits | DSP Blocks | Pir
1 |SoCKit_DDR3_RTL_Test 87624 (154) 5982 (135) 1540512 111 242
1 |ALU:alu| 31996 (703) 869 (869) 0 111 0
1 |backprop_pool:bp| 100 (100) 0 (0) 0 0 0
2 |convolution:convo| 31044 (0) 0 (0) 0 95 0
3 |dot_product:dp| 149 (149) 0 (0) 0 16 0
2 |fpga_ddr3:fpga_ddr3_inst| 4181 (0) 3951 (0) 213408 0 0
3 |mem_control:mv| 51198 (2093) 950 (0) 1327104 0 0
1 |mask_buffer:mb_i| 58 (58) 288 (288) 0 0 0
2 |read_buffer:rb1_i| 17636 (83) 75 (66) 442368 0 0
3 |read_buffer:rb2_i| 16161 (82) 66 (66) 442368 0 0
4 |write_back_accumulator:wba_i| 15250 (105) 521 (69) 442368 0 0
4 |sld_hub:auto_hub| 95 (1) 77 (0) 0 0 0
Note: For table entries with two numbers listed, the numbers in parentheses indicate the number of resources of the given type used by the specific entity alone. The numbers listed outside of parentheses indicate the total resources of the given type used by the specific entity and all of its sub-entities in the hierarchy.
IP Catalog
Library
Basic Functions
Arithmetic
Bridges and Ad
Clocks; PLLs a
Configuration a
I/O
Miscellaneous
On Chip Memo
FIFO
RAM ini
RAM: 1-
RAM: 2-
ROM: 1
ROM: 2
Shift reg
Simulation; Deb
Debug and
Simulation
Verification
DSP
Interface Protocols
Memory Interfaces
Add...
Type ID Message
128002 Starting physical synthesis algorithm register retiming
128003 Physical synthesis algorithm register retiming complete: estimated slack improvement of 0 ps
128002 Starting physical synthesis algorithm register retiming
128003 Physical synthesis algorithm register retiming complete: estimated slack improvement of 0 ps
128001 Physical synthesis optimizations for speed complete: elapsed time is 00:02:54
176233 Starting register packing
176235 Finished register packing
176219 No registers were packed into other blocks
15705 Ignored locations or region assignments to the following nodes
11798 Fitter preparation operations ending: elapsed time is 00:05:35
170189 Fitter placement preparation operations beginning
170011 Design contains 87624 blocks of type combinational node. However, the device contains only 83820 blocks.
14951 The Fitter is using Advanced Physical Optimization.
Messages
System
Processing (1142)