Lab 5 report
Yanzhe Liu 504153440

1. Parallelization Strategy
I used Merlin compiler for this lab. I was trying to mimic the coding style from the "Optimizing FPGA-Based Accelerator Design for Deep Convolutional Neural Networks" paper. Unfortunately, I have little control over how Merlin create the local buffer. So I decided to only tile the h&w loop. When tiled hh and ww loops are the innermost loops, merlin would perform reduction technique to help speed up the process.

```
h0 loop
  w0 loop
    q loop
      p loop
     i loop
     #pragma ACCEL pipeline
       j loop
         #pragma ACCEL parallel flatten
         hh loop
           ww loop
```

2. The execution time for lab5 is significantly lower than lab3 and lab4.

	Performance	Power	Efficiency
lab3:	120 GFlops	95 W	1.26315789
lab4:	70 GFlops	234 W	0.29914529914
lab5:	5 GFlops	25 W	0.2

We can see that CPU is the most power efficient device. FPGA and GPU efficiencis are about the same. However, if I have a better FPGA optimization, it could have higher efficiency than GPU.

3.
The main challenge of this lab is to learn how to utilize the built-in function of merlin and understanding the resource limitation of FPGA.