

# Program output

|             | Checksum           | Time in millesecond |
|-------------|--------------------|---------------------|
| naive cuda  | 122756344698240.00 | 0.038               |
| tilted cuda | 122756344698240.00 | 0.017               |
| cuDNN       | 122756344698240.00 | 0.082               |

## How to run

```
module purge;

module load cuda/9.0.176;

module load cudnn/9.0v7.0.5

make clean && make

sbatch run.sh
```

result is in lab3.out

## Discussion

Why cuDNN is even slower? I think it is because cuDNN might preform some pre-process work like padding and striding when running **cudnnGetConvolutionForwardAlgorithm** function, which means it is not a pure kernel call.

Also, using shared memory speeds up convolution computation.

If we can put filter data into constant memory in gpu, I believe the speed up is far more drastic. (The only question is whether large filter kernel with large K and FW/FH can suit constant area or not)

