UNIVERSITÉ DE PAU ET DES PAYS DE L'ADOUR

M2-BIG DATA
GPGPU - Chapter 8

Exercice 1

Réalisé par:
Gaby Maroun

Encadré par:
Dr. Etancelin JM

February 6, 2021

Objectives

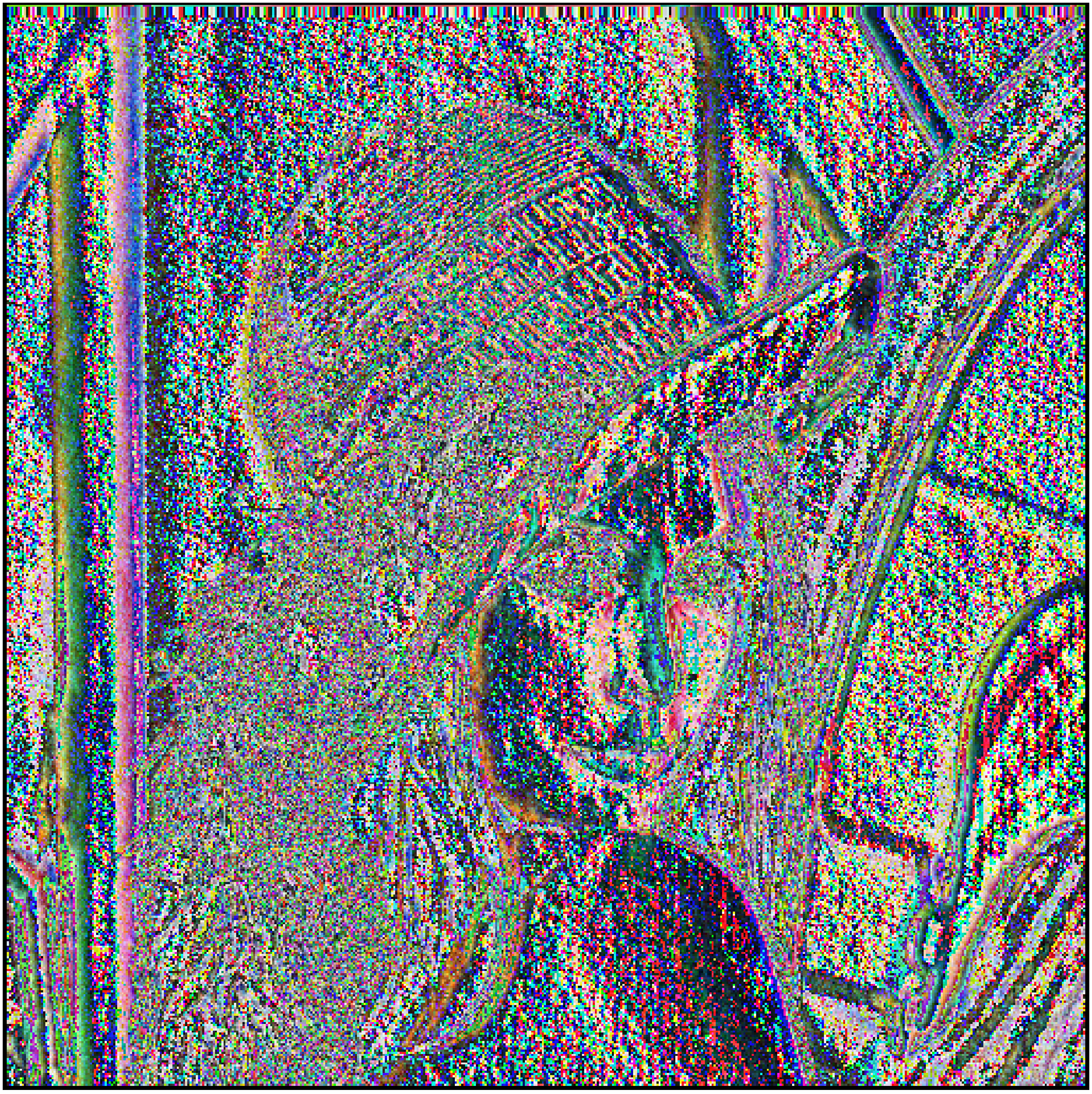
Implement a convolution filter for constant squared boxed filters. The convolution formula is given by :

$$P_{i,j,c} = \sum_{x=-k}^{k} \sum_{y=-k}^{k} I_{i+x,j+y,c} M_{x,y}$$

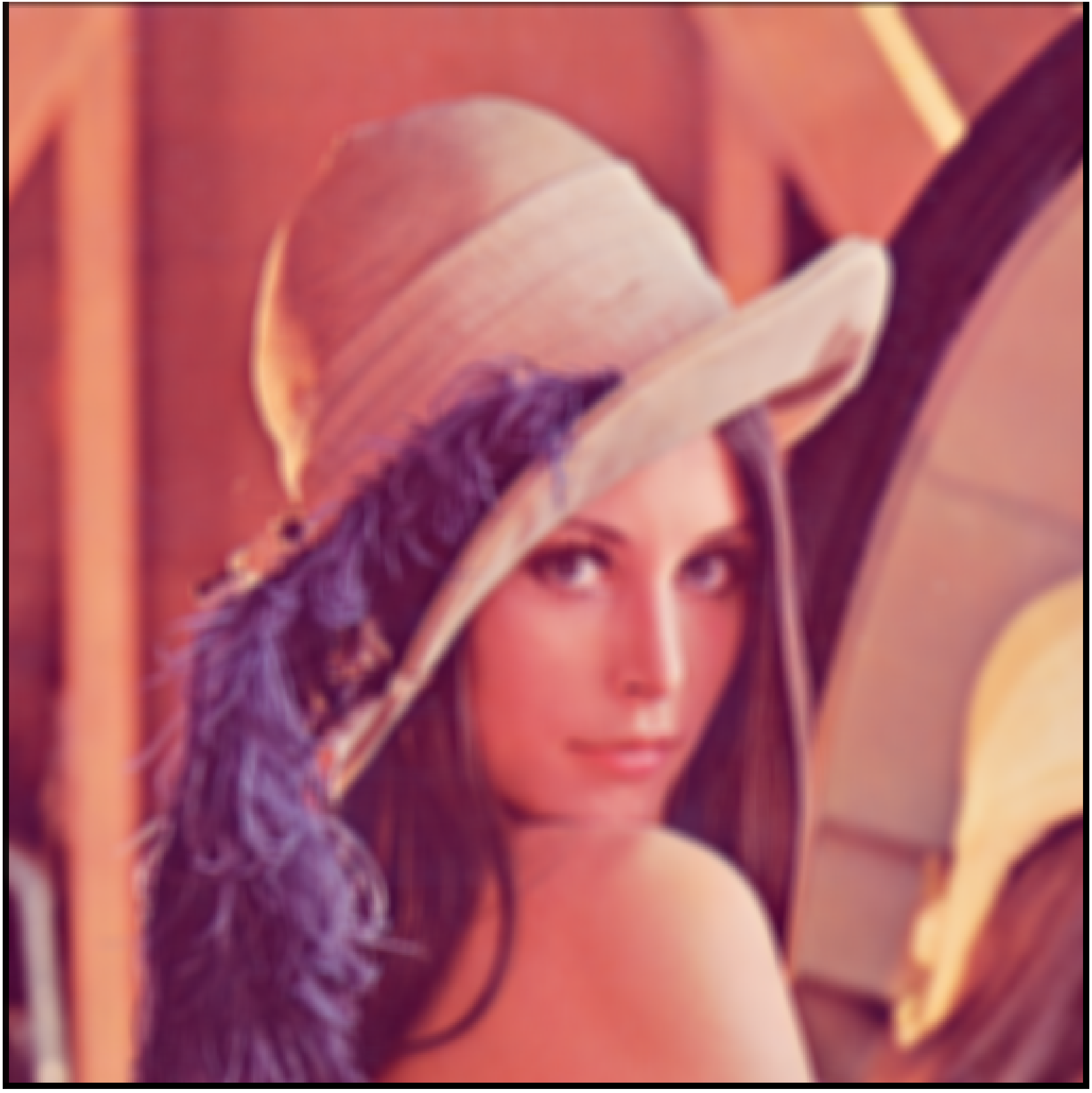
where c is image channel, M is the filter mask of size $2k+1$ the parameter k is called filter radius.

Instructions

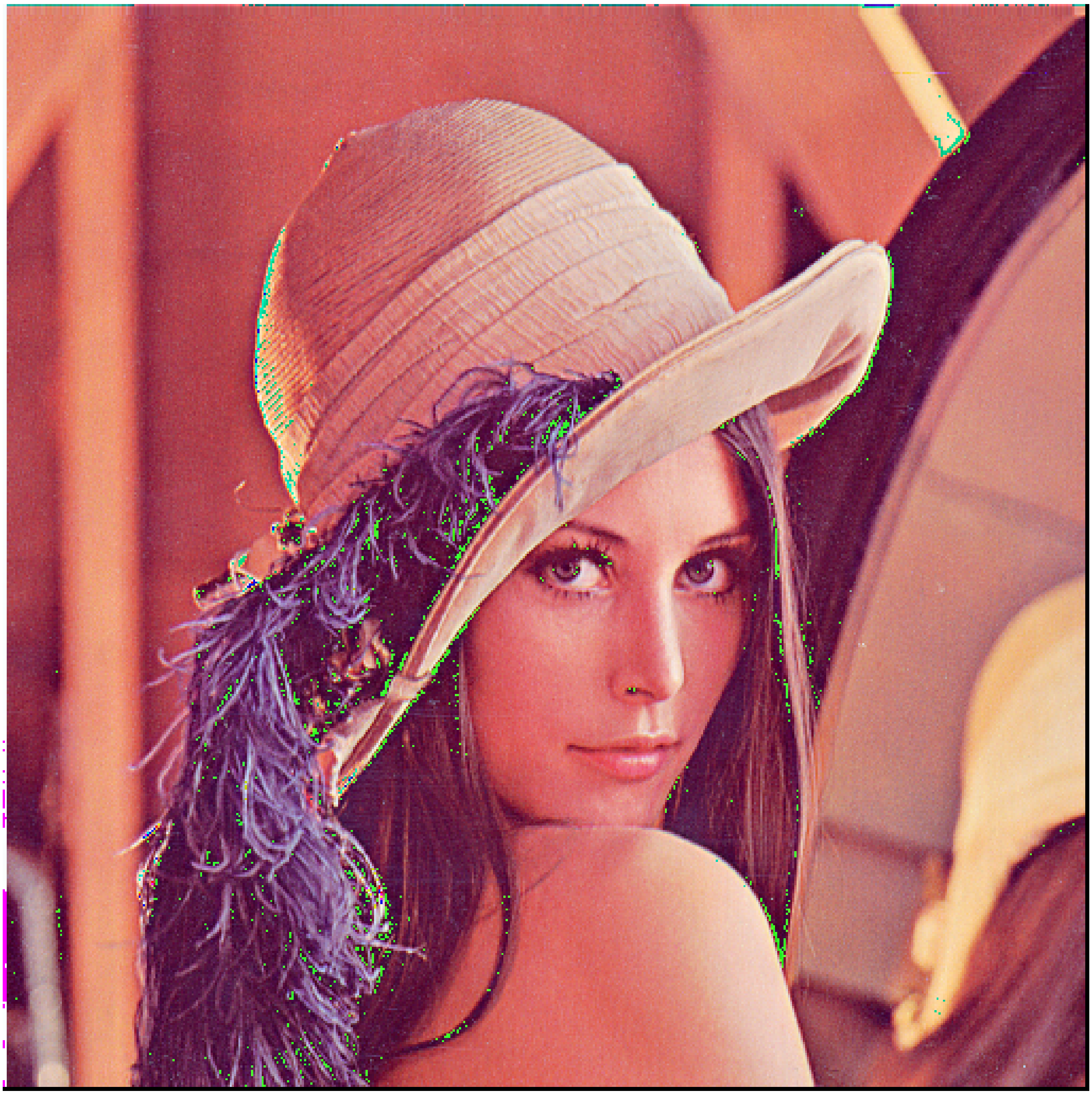
- $k = 1$ and $M = \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$



- $k = 1$ and $M = (\frac{1}{25})$



- $k = 1$ and $M = (\frac{-1}{256})\begin{pmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & -476 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{pmatrix}$



Questions

1. **How many floating operations are being performed in your convolution kernel ? explain.**

   There are

   $$imgCols \times imgRows \times channels(2 \times maskWidth^2)$$

   *floating operations for cycling on all of the image with all the numbers of channels and also cycling through the mask twice, one for + and the other to **

2. **How many global memory reads are being performed by your kernel ? explain.**

   There are

   $$imgCols \times imgRows \times channels(2 \times maskWidth^2)$$

*global memory reads for cycling to read from all the pixels of the image for the different number of channels and also cycling through the mask twice, one for + and the other to **

3. **How many global memory writes are being performed by your kernel ? explain.**
There are

$$imgCols \times imgRows \times channels$$

global memory writes for writing the image for different number of channels

4. **Compute the arithmetic intensity of the kernel.**
The arithmetic intensity is a FLOP/Byte number standing for the number of floating point operations performed per byte of global memory accessed.

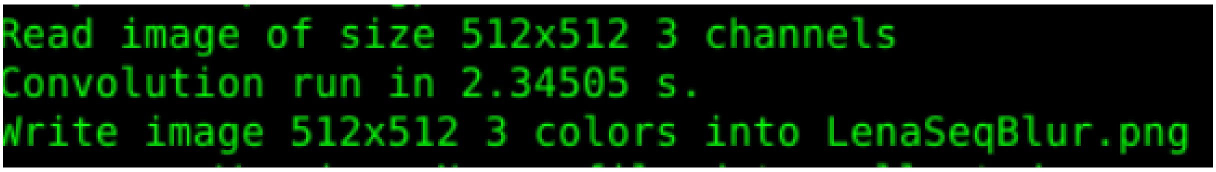
$$\frac{imgCols \times imgRows \times channels(2 \times maskWidth^2)}{imgCols \times imgRows \times channels(2 \times maskWidth^2) + imgCols \times imgRows \times channels}$$

$$=> \frac{1}{1 + \frac{imgCols \times imgRows \times channels}{imgCols \times imgRows \times channels(2 \times maskWidth^2)}} (FLOP/Byte)$$

$$=> \frac{1}{1 + \frac{1}{(2 \times maskWidth^2)}} (FLOP/Byte)$$

5. **Measure the kernel computational time of the kernel, using the profiler. Then, compute the computational power of the kernel (in GFLOPS). Compare with the CPU version given.**

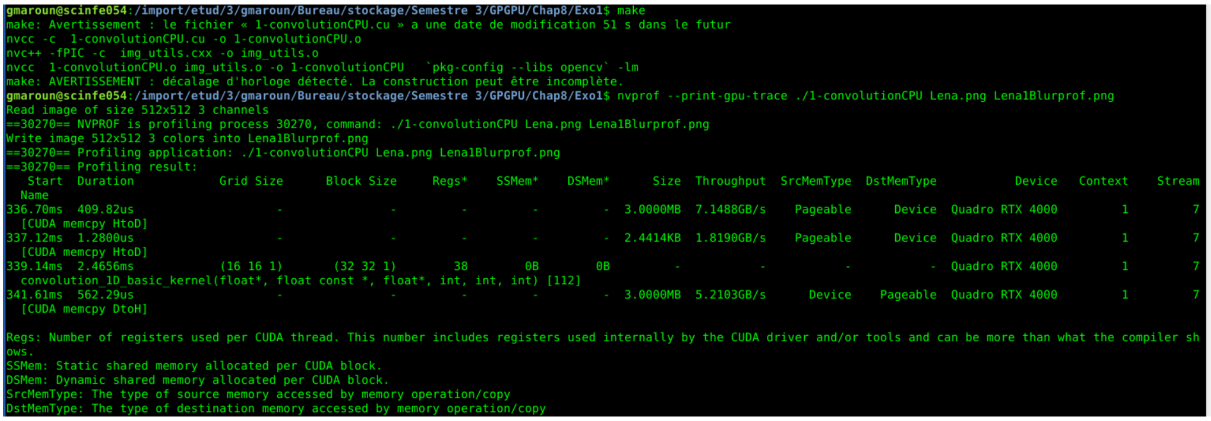
Sequential Version:



```
Read image of size 512x512 3 channels
Convolution run in 2.34505 s.
Write image 512x512 3 colors into LenaSeqBlur.png
```

*As it seems, it takes 2.34505seconds for the kernel to compile with a 25*25 mask on a 512x512 image size.*

Parallel version:



```
gmaroun@scinfe054:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap8/Exo1$ make
make: Avertissement : le fichier « 1-convolutionCPU.cu » a une date de modification 51 s dans le futur
nvcc -c  1-convolutionCPU.cu -o 1-convolutionCPU.o
nvc++ -fPIC -c  img_utils.cxx -o img_utils.o
nvcc  1-convolutionCPU.o img_utils.o -o 1-convolutionCPU   `pkg-config --libs opencv` -lm
make: AVERTISSEMENT : décalage d'horloge détecté. La construction peut être incomplète.
gmaroun@scinfe054:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap8/Exo1$ nvprof --print-gpu-trace ./1-convolutionCPU Lena.png Lena1Blurprof.png
Read image of size 512x512 3 channels
==30270== NVPROF is profiling process 30270, command: ./1-convolutionCPU Lena.png Lena1Blurprof.png
Write image 512x512 3 colors into Lena1Blurprof.png
==30270== Profiling application: ./1-convolutionCPU Lena.png Lena1Blurprof.png
==30270== Profiling result:
   Start  Duration            Grid Size      Block Size     Regs*    SSMem*    DSMem*      Size  Throughput  SrcMemType  DstMemType           Device   Context    Stream
  Name
336.70ms  409.82us                    -               -         -         -         -  3.0000MB  7.1488GB/s    Pageable      Device  Quadro RTX 4000         1         7
  [CUDA memcpy HtoD]
337.12ms  1.2800us                    -               -         -         -         -  2.4414KB  1.8190GB/s    Pageable      Device  Quadro RTX 4000         1         7
  [CUDA memcpy HtoD]
339.14ms  2.4656ms            (16 16 1)       (32 32 1)        38        0B        0B         -           -           -           -  Quadro RTX 4000         1         7
  convolution_1D_basic_kernel(float*, float const *, float*, int, int, int) [112]
341.61ms  562.29us                    -               -         -         -         -  3.0000MB  5.2103GB/s      Device    Pageable  Quadro RTX 4000         1         7
  [CUDA memcpy DtoH]

Regs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler sh
ows.
SSMem: Static shared memory allocated per CUDA block.
DSMem: Dynamic shared memory allocated per CUDA block.
SrcMemType: The type of source memory accessed by memory operation/copy
DstMemType: The type of destination memory accessed by memory operation/copy
```

*As it seems, it takes 2.4656milliseconds for the kernel to compile with a 25*25 mask on a 512x512 image size.*

For the computational power of the kernel for the GPU version, it is equal to :

$$\frac{FloatingOperations}{ExecutionTime}$$

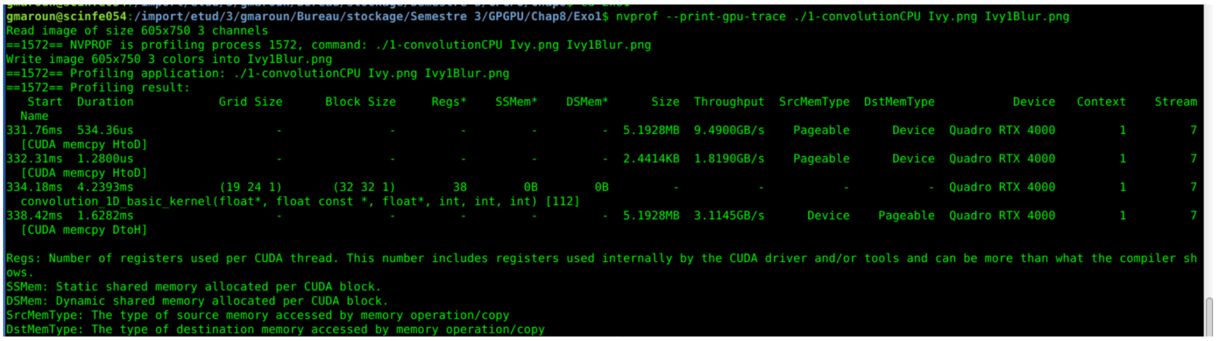
$$=> \frac{imgCols \times imgRows \times channels(2 \times maskWidth^2)}{2.4656ms}$$

For the computational power of the kernel for the CPU version and having the same number of floating operations, so :

$$\frac{FloatingOperations}{2.4656 \times 10^3 s} < \frac{FloatingOperations}{2.34505s}$$

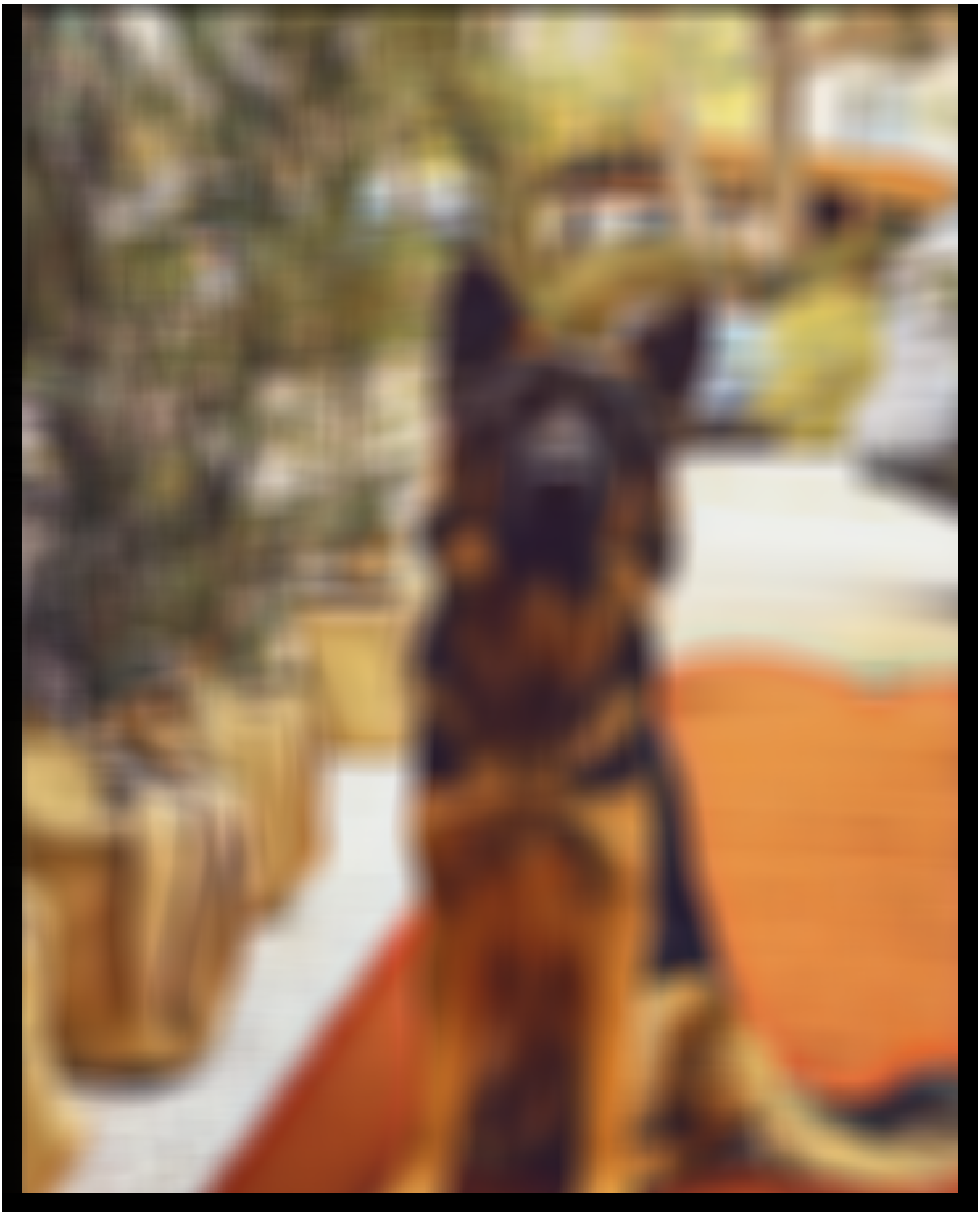
That means, this exercise's kernel has 10^3 more computation power than the CPU's

6. **Compare the computational power evolution using different images sizes.**



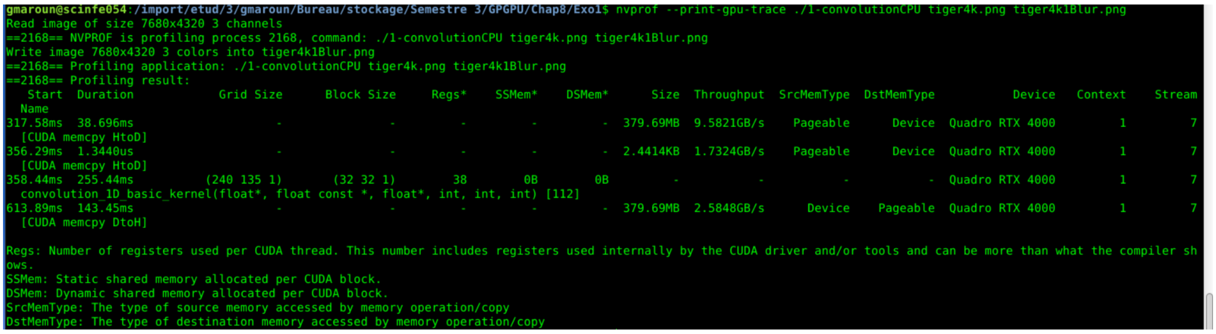
```
gmaroun@scinfe054:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap8/Exo1$ nvprof --print-gpu-trace ./1-convolutionCPU Ivy.png Ivy1Blur.png
Read image of size 605x750 3 channels
==1572== NVPROF is profiling process 1572, command: ./1-convolutionCPU Ivy.png Ivy1Blur.png
Write image 605x750 3 colors into Ivy1Blur.png
==1572== Profiling application: ./1-convolutionCPU Ivy.png Ivy1Blur.png
==1572== Profiling result:
   Start  Duration            Grid Size      Block Size     Regs*    SSMem*    DSMem*      Size  Throughput  SrcMemType  DstMemType           Device   Context    Stream
  Name
331.76ms  534.36us                    -               -         -         -         -  5.1928MB  9.4900GB/s    Pageable      Device  Quadro RTX 4000         1         7
  [CUDA memcpy HtoD]
332.31ms  1.2800us                    -               -         -         -         -  2.4414KB  1.8190GB/s    Pageable      Device  Quadro RTX 4000         1         7
  [CUDA memcpy HtoD]
334.18ms  4.2393ms            (19 24 1)       (32 32 1)        38        0B        0B         -           -           -           -  Quadro RTX 4000         1         7
  convolution_1D_basic_kernel(float*, float const *, float*, int, int, int) [112]
338.42ms  1.6282ms                    -               -         -         -         -  5.1928MB  3.1145GB/s      Device    Pageable  Quadro RTX 4000         1         7
  [CUDA memcpy DtoH]

Regs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler sh
ows.
SSMem: Static shared memory allocated per CUDA block.
DSMem: Dynamic shared memory allocated per CUDA block.
SrcMemType: The type of source memory accessed by memory operation/copy
DstMemType: The type of destination memory accessed by memory operation/copy
```

Ivy :



*it takes 4.2393 milliseconds for the kernel to compile with a 25*25 mask on a 605x750 image size*

Tiger4K :



```
gmaroun@scinfe054:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap8/Exo1$ nvprof --print-gpu-trace ./1-convolutionCPU tiger4k.png tiger4k1Blur.png
Read image of size 7680x4320 3 channels
==2168== NVPROF is profiling process 2168, command: ./1-convolutionCPU tiger4k.png tiger4k1Blur.png
Write image 7680x4320 3 colors into tiger4k1Blur.png
==2168== Profiling application: ./1-convolutionCPU tiger4k.png tiger4k1Blur.png
==2168== Profiling result:
   Start  Duration            Grid Size      Block Size     Regs*    SSMem*    DSMem*      Size  Throughput  SrcMemType  DstMemType           Device   Context    Stream
  Name
317.58ms  38.696ms                    -               -         -         -         -  379.69MB  9.5821GB/s    Pageable      Device  Quadro RTX 4000         1         7
  [CUDA memcpy HtoD]
356.29ms  1.3440us                    -               -         -         -         -  2.4414KB  1.7324GB/s    Pageable      Device  Quadro RTX 4000         1         7
  [CUDA memcpy HtoD]
358.44ms  255.44ms          (240 135 1)       (32 32 1)        38        0B        0B         -           -           -           -  Quadro RTX 4000         1         7
  convolution_1D_basic_kernel(float*, float const *, float*, int, int, int) [112]
613.89ms  143.45ms                    -               -         -         -         -  379.69MB  2.5848GB/s      Device    Pageable  Quadro RTX 4000         1         7
  [CUDA memcpy DtoH]

Regs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler sh
ows.
SSMem: Static shared memory allocated per CUDA block.
DSMem: Dynamic shared memory allocated per CUDA block.
SrcMemType: The type of source memory accessed by memory operation/copy
DstMemType: The type of destination memory accessed by memory operation/copy
```

*it takes 255.44milliseconds for the kernel to compile with a 25*25 mask on a 7680x4320 image size*

La fin.