

Lập trình song song và ứng dụng

Enhancing Image Document

20120322 : Đặng Thiên Long
20120510 : Nguyễn Hữu Anh Khoa

01 Tổng Quan

02 Phiên Bản Tối Ưu Lần 1

03 Phiên Bản Tối Ưu Lần 2

04 Phiên Bản Tối Ưu Lần 3

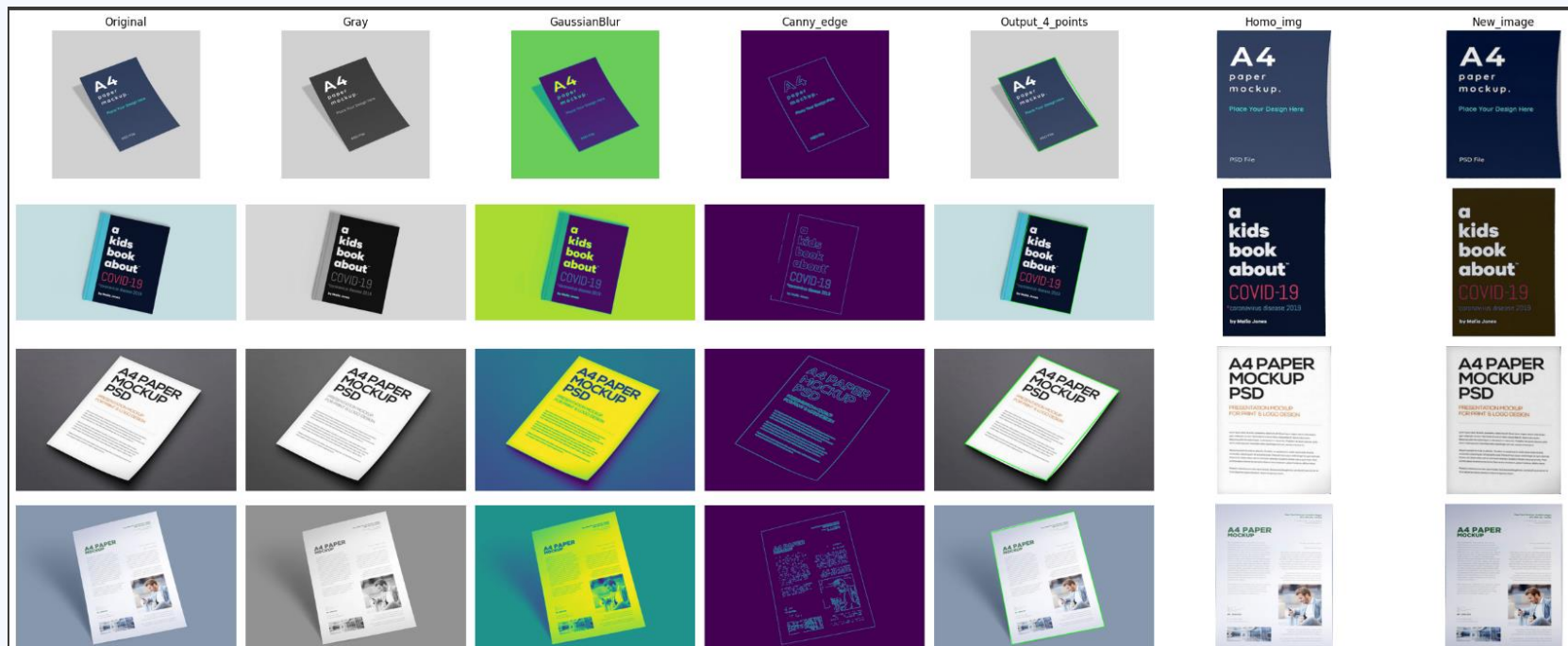
05 Tổng Kết

01

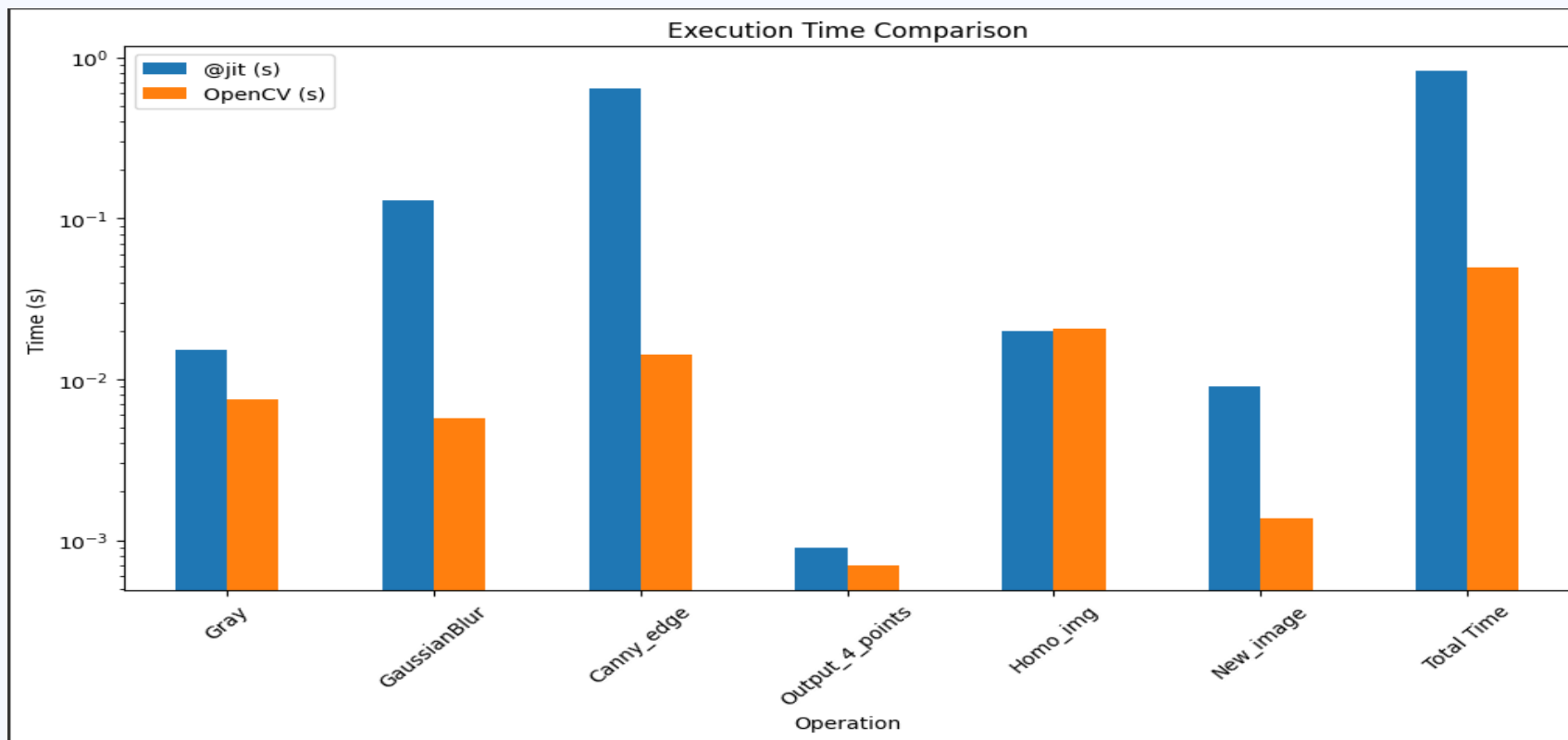
Tổng Quan



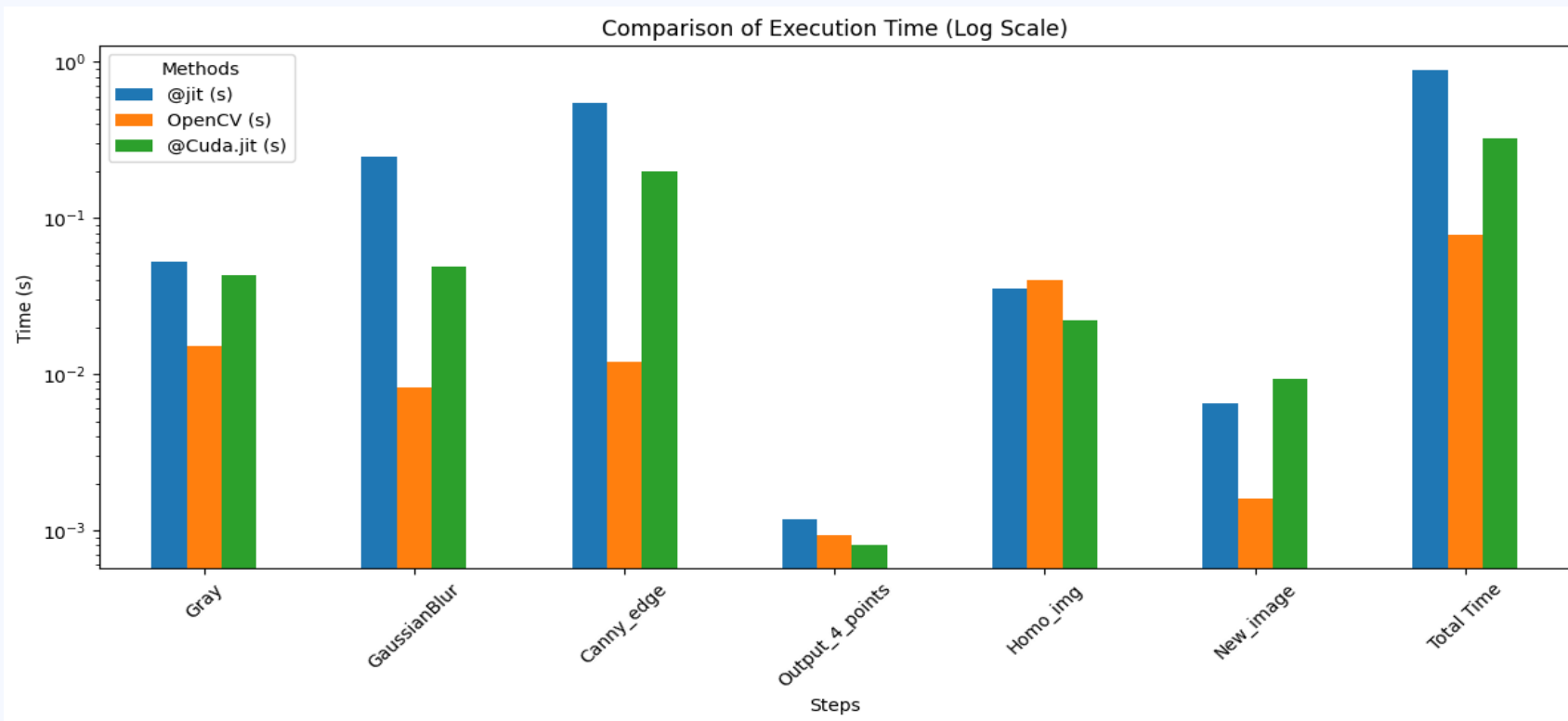
Tổng quan



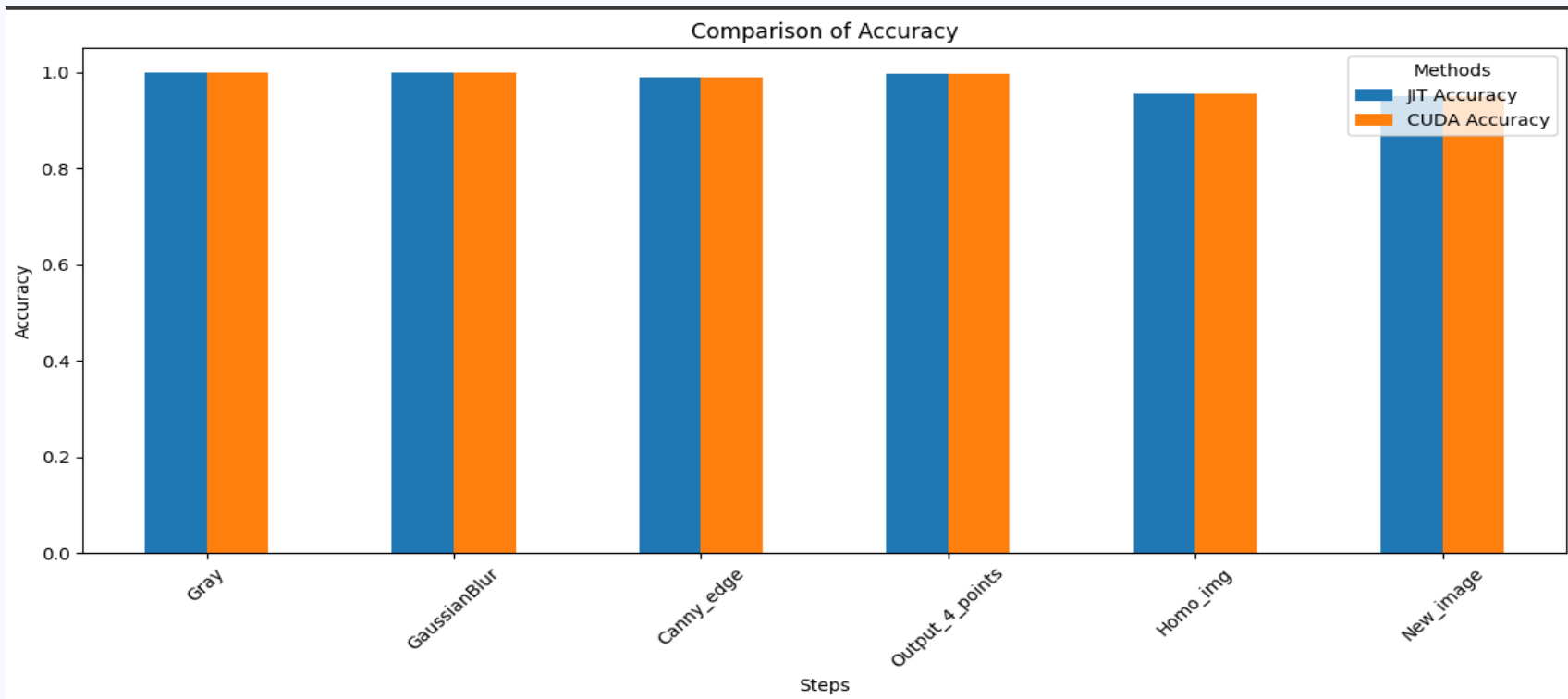
Tổng quan



Tổng quan



Tổng quan



02

Phiên bản tối ưu lần 1



Phiên bản tối ưu lần 1

```
--34024-- NvPROF is profiling process 34024, command: python3 parallel.py --config config.json

Processed time: 7.651118993759155 s
--34024-- Profiling application: python3 parallel.py --config config.json
--34024-- Profiling result:

  Start   Duration      Grid Size      Block Size      Regs*      SSMem*      Size      Throughput      SrcMemType      DstMemType      Device      Context      Stream      Name
563.89ms  184.67us      -              -              -          -          - 10.466GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
564.70ms  596.75us      -              -              -          -          - 2.9898MB 4.8927GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
565.37ms  94.045us      (20 20 1)      (32 32 1)      17          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
565.52ms  183.17us      -              -              -          -          - 1.1212MB 10.613GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
565.98ms  424.02us      -              -              -          -          - 2.9898MB 6.8858GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
569.19ms  318.26us      -              -              -          -          - 2.1629MB 6.6368GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
570.33ms  1.6823ms      -              -              -          -          - 5.7678MB 3.3481GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
572.06ms  179.68us      (38 20 1)      (32 32 1)      17          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
572.24ms  266.26us      -              -              -          -          - 2.1629MB 7.9329GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
572.83ms  2.1471ms      -              -              -          -          - 5.7678MB 2.6234GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
577.84ms  111.23us      -              -              -          -          - 1.1856MB 10.409GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
578.69ms  794.57us      -              -              -          -          - 3.1616MB 3.8858GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
579.50ms  98.589us      (25 17 1)      (32 32 1)      17          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
579.60ms  189.34us      -              -              -          -          - 1.1856MB 10.589GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
580.02ms  2.1472ms      -              -              -          -          - 3.1616MB 1.4379GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
584.72ms  289.56us      -              -              -          -          - 1.8797MB 6.3394GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
585.82ms  1.5433ms      -              -              -          -          - 5.0125MB 3.1718GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
587.37ms  158.43us      (32 21 1)      (32 32 1)      17          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
587.53ms  166.88us      -              -              -          -          - 1.8797MB 11.006GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
588.09ms  4.5317ms      -              -              -          -          - 5.0125MB 1.0802GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
594.64ms  67.838us      -              -              -          -          - 766.61KB 10.777GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
595.50ms  491.70us      -              -              -          -          - 1.9964MB 3.9650GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
596.00ms  65.631us      (20 14 1)      (32 32 1)      17          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
596.10ms  64.686us      -              -              -          -          - 766.61KB 11.316GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
596.42ms  176.25us      -              -              -          -          - 1.1646MB 11.061GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
601.59ms  1.0001ms      -              -              -          -          - 5.4840MB 5.3549GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
603.50ms  5.4429ms      -              -              -          -          - 14.624MB 2.6239GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
609.02ms  447.12us      (50 38 1)      (32 32 1)      17          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
609.47ms  837.71us      -              -              -          -          - 5.4840MB 6.3938GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
610.66ms  13.938ms      -              -              -          -          - 14.624MB 1.0247GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
703.12ms  533.62us      -              -              -          -          - 2.9898MB 5.4215GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
704.43ms  577.29us      -              -              -          -          - 2.9898MB 5.0576GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
706.67ms  902ns        -              -              -          -          - 728 69.218MB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
706.79ms  158.68us      (20 20 1)      (32 32 1)      58          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
706.95ms  433.24us      -              -              -          -          - 2.9898MB 6.7393GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
707.91ms  394.97us      -              -              -          -          - 2.9898MB 7.3923GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
708.77ms  1.8880ms      -              -              -          -          - 728 36.369MB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
709.60ms  1.1275ms      -              -              -          -          - 5.7678MB 4.9956GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
801.40ms  1.5749ms      -              -              -          -          - 5.7678MB 3.5766GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
803.27ms  736ns        -              -              -          -          - 728 93.294MB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
803.33ms  296.89us      (38 20 1)      (32 32 1)      58          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
803.63ms  1.0335ms      -              -              -          -          - 5.7678MB 5.4502GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
805.14ms  2.1653ms      -              -              -          -          - 5.7678MB 2.6013GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
809.22ms  2.1440ms      -              -              -          -          - 728 32.026MB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
810.09ms  547.25us      -              -              -          -          - 3.1616MB 5.6419GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
811.48ms  856.04us      -              -              -          -          - 3.1616MB 3.6038GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
812.66ms  672ns        -              -              -          -          - 728 102.18MB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
812.71ms  164.60us      (25 17 1)      (32 32 1)      58          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
812.88ms  580.85us      -              -              -          -          - 3.1616MB 5.3155GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
813.89ms  2.1918ms      -              -              -          -          - 3.1616MB 1.4087GB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
817.43ms  2.1440ms      -              -              -          -          - 728 32.026MB/s      Device      Pageable      Tesla T4 (0)  1          7          [CUDA memcopy Dtoh]
818.22ms  1.0124ms      -              -              -          -          - 5.0125MB 4.8350GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
820.14ms  1.5158ms      -              -              -          -          - 5.0125MB 3.2293GB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
821.95ms  672ns        -              -              -          -          - 728 102.18MB/s      Pageable      Device      Tesla T4 (0)  1          7          [CUDA memcopy HtoD]
822.08ms  262.55us      (32 21 1)      (32 32 1)      58          0B          0B -              -          -          Tesla T4 (0)  1          7          _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUhw1sCUT9Uw11Ew1dRRQPKZ
```

Phiên bản tối ưu lần 1

- Có thể thấy sao chép dữ liệu từ CPU sang GPU và ngược lại tốn nhiều thời gian nhất.
- Ở phiên bản song song hiện tại thì dữ liệu sẽ được copy sang GPU mỗi khi gọi hàm và copy về CPU sau khi thực hiện xong hàm đó
=> chỉ copy những dữ liệu cần thiết, hạn chế copy qua lại nhiều lần giữa CPU và GPU.

Phiên bản tối ưu lần 1

```
# Chuyển dữ liệu ảnh từ CPU sang GPU
```

```
from numba import cuda
```

```
d_img = [cuda.to_device(img[i]) for i in range(len(img))]
```

```
d_gray_op1 = [cuda.device_array((d_img[i].shape[0], d_img[i].shape[1]), img[i].dtype) for i in range(len(img))]
```

```
grid_size = [(math.ceil(d_gray_op1[i].shape[1] / block_size[0]),  
              math.ceil(d_gray_op1[i].shape[0] / block_size[1])) for i in range(len(img))]
```

```
for i in range(len(img)):  
    convert_RGB_To_Gray_cuda[grid_size[i], block_size](d_img[i], d_gray_op1[i])
```

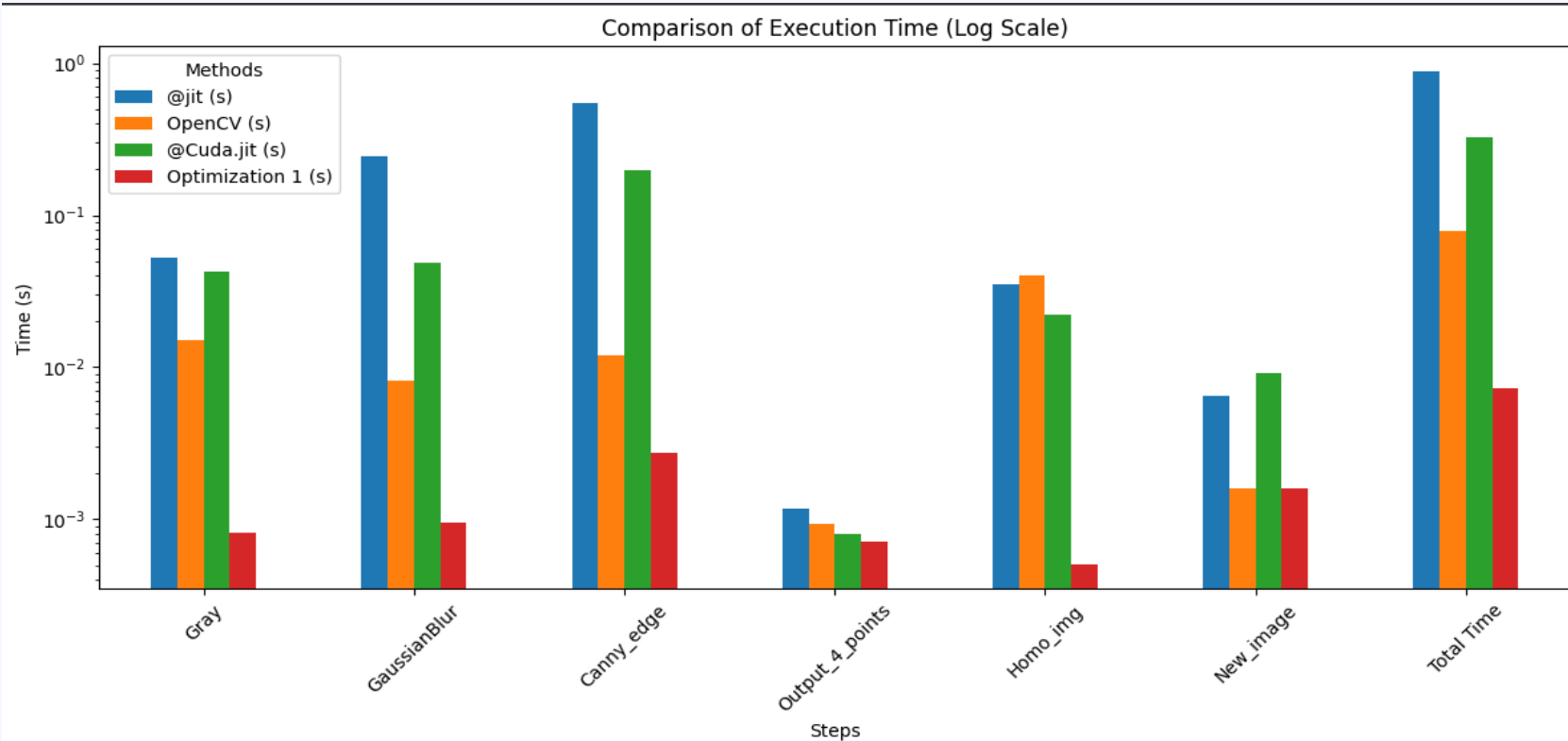
```
start_time = time.time()
```

```
for i in range(len(img)):  
    convert_RGB_To_Gray_cuda[grid_size[i], block_size](d_img[i], d_gray_op1[i])  
end_time = time.time()
```

```
execution_gray_op1_time = end_time - start_time
```

```
gray_op1 = [d_gray_op1[i].copy_to_host() for i in range(len(img))]
```

Phiên bản tối ưu lần 1



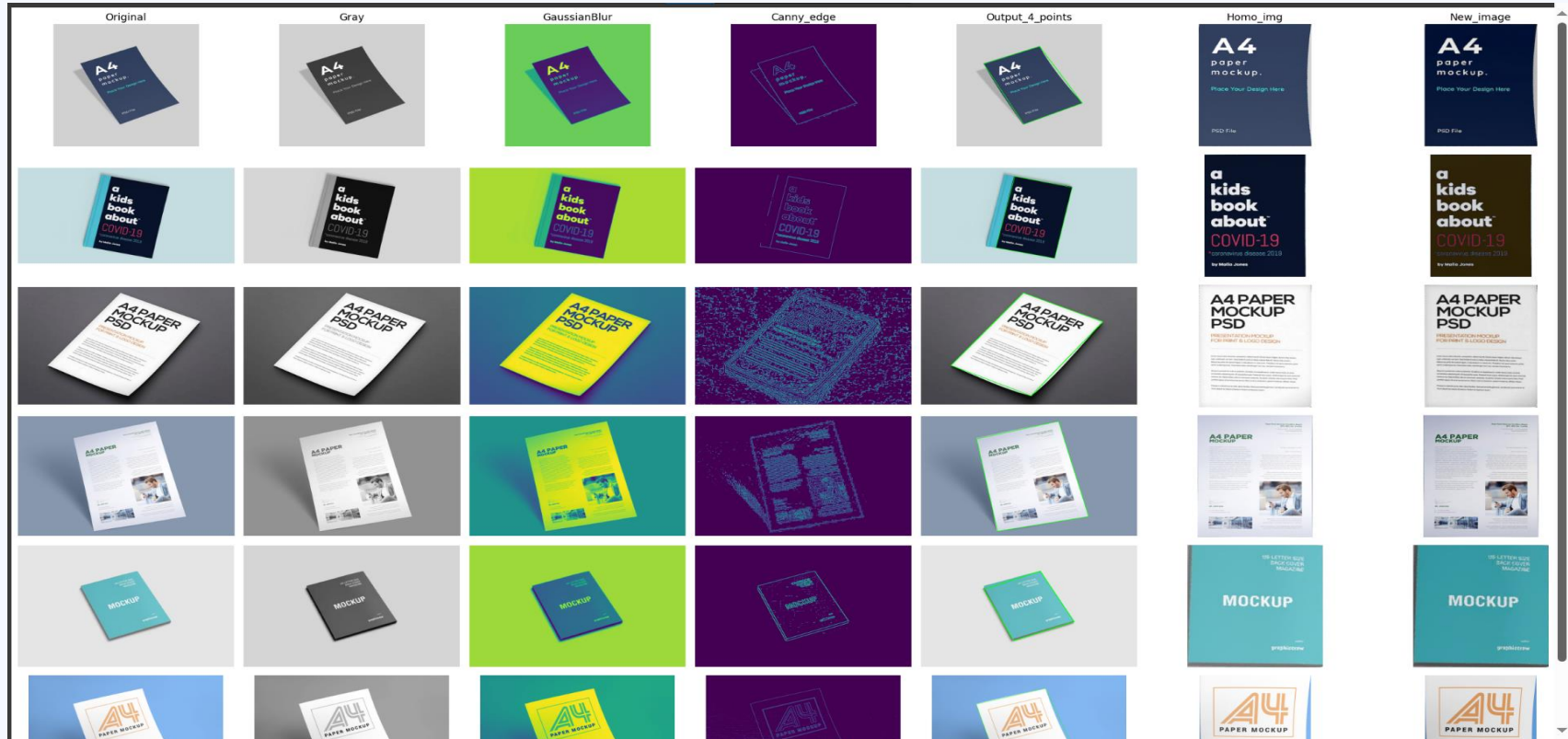
Phiên bản tối ưu lần 1

	@jit (s)	OpenCV (s)	@Cuda.jit (s)	Optimization 1 (s)
Gray	0.020392	0.006911	0.053736	0.000551
GaussianBlur	0.122405	0.005718	0.046141	0.001040
Canny_edge	0.572253	0.011081	0.228040	0.002144
Output_4_points	0.000954	0.001349	0.003944	0.000795
Homo_img	0.035583	0.034992	0.027035	0.000863
New_image	0.005895	0.005548	0.006176	0.001513
Total Time	0.757482	0.065599	0.365071	0.006907

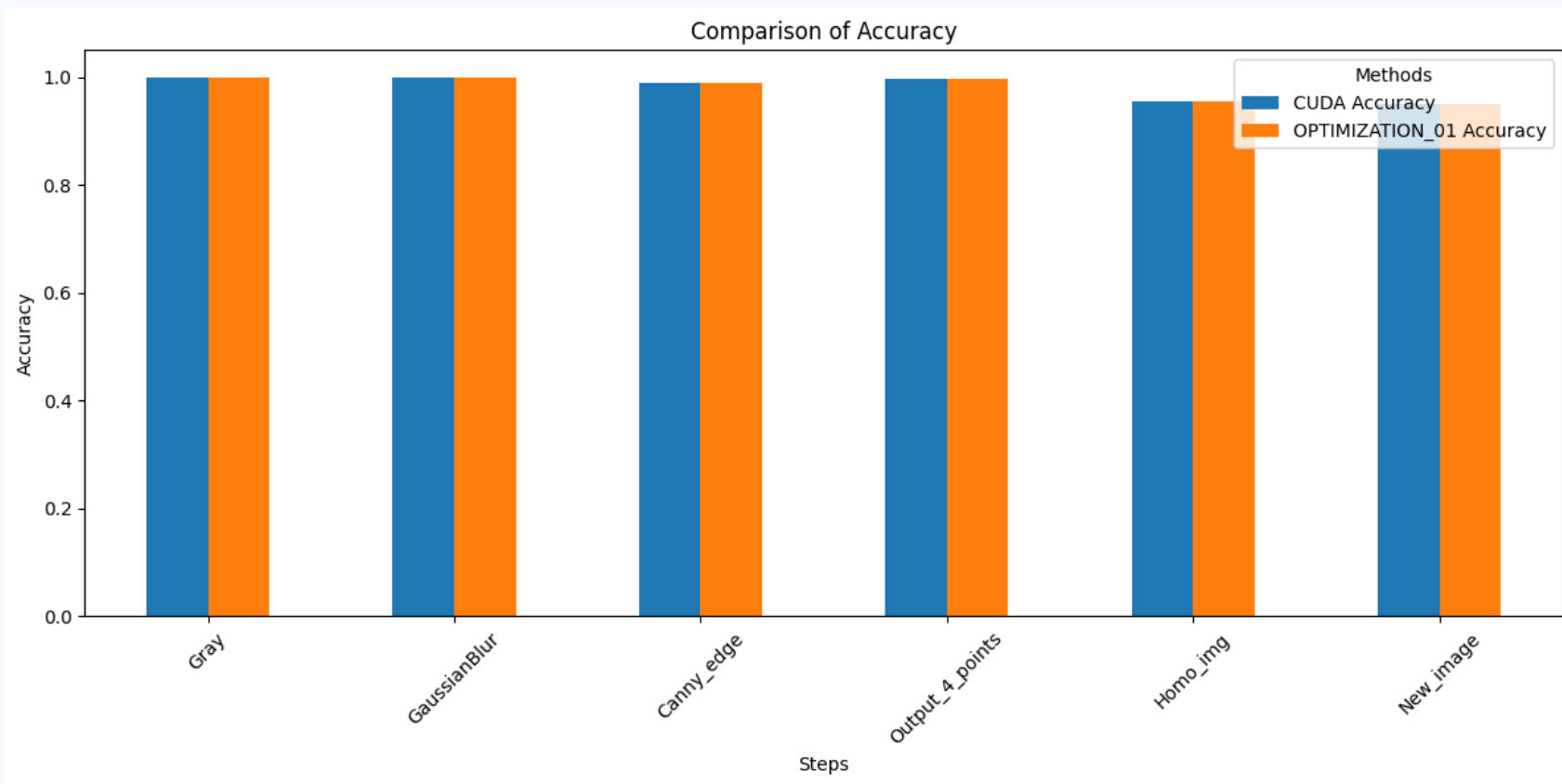
Phiên bản tối ưu lần 1

- Gray: Optimization 1 nhanh nhất, nhanh hơn @jit khoảng 34.9 lần , OpenCV khoảng 49.4 lần, và @Cuda.jit khoảng 46.4 lần .
- GaussianBlur: Optimization 1 nhanh nhất, nhanh hơn @jit khoảng 460 lần, OpenCV khoảng 31.9 lần, và @Cuda.jit khoảng 84 lần.
- Canny_edge: Optimization 1 nhanh nhất, nhanh hơn @jit khoảng 112.4 lần , OpenCV khoảng 6.9 lần , và @Cuda.jit khoảng 62 lần.
- Homo_img: Optimization 1 nhanh nhất, nhanh hơn @jit khoảng 35.4 lần, OpenCV khoảng 36.2 lần , và @Cuda.jit khoảng 41.4 lần.
- Tổng thời gian : Optimization 1 nhanh nhất, nhanh hơn @jit khoảng 82 lần, nhanh hơn OpenCV khoảng 12.2 lần , và nhanh hơn @Cuda.jit khoảng 38.9 lần.

Phiên bản tối ưu lần 1

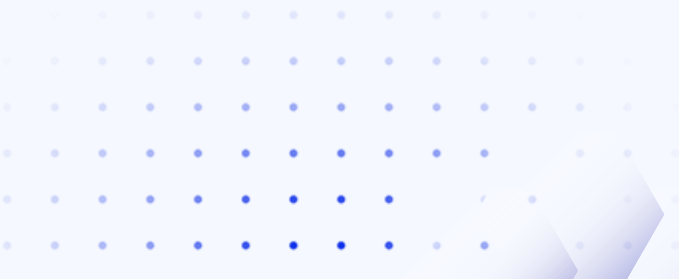


Phiên bản tối ưu lần 1



03

Phiên bản tối ưu lần 2



Phiên bản tối ưu lần 2

```
--911-- NVPROF is profiling process 911, command: python3 parallel_optimize_01.py --config config-json
--911-- Profiling application: python3 parallel_optimize_01.py --config config-json
--911-- Profiling result:
Start Duration Grid Size Block Size Regs* SSMem* DSMem* Size Throughput SrcMemType DstMemType Device Context Stream Name
218.44ms 101.31us - - - - - 1.1212MB 10.887GB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
219.12ms 327.61us - - - - - 2.1629MB 6.4475GB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
221.23ms 105.92us - - - - - 1.1856MB 10.931GB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
223.12ms 275.77us - - - - - 1.8797MB 6.6565GB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
223.87ms 66.302us - - - - - 766.64KB 11.027GB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
226.41ms 951.37us - - - - - 5.4848MB 5.6292GB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
510.68ms 88.318us (20 20 1) (32 32 1) 17 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ewd
510.82ms 165.28us (38 20 1) (32 32 1) 17 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ewd
510.90ms 91.806us (25 17 1) (32 32 1) 17 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ewd
511.00ms 146.56us (32 21 1) (32 32 1) 17 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ewd
511.23ms 62.654us (20 14 1) (32 32 1) 17 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ewd
511.29ms 411.19us (50 38 1) (32 32 1) 17 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ewd
511.71ms 32.384us - - - - - - 382.69KB 11.270GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
511.92ms 60.287us - - - - - - 738.28KB 11.679GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
512.21ms 33.400us - - - - - - 404.69KB 11.552GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
512.30ms 52.062us - - - - - - 641.60KB 11.753GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
512.63ms 22.016us - - - - - - 255.54KB 11.069GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
512.75ms 150.17us - - - - - - 1.8280MB 11.887GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
515.17ms 1.6320us (20 20 1) (32 32 1) 62 0B 0B 72B 42.074MB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
720.53ms 220.74us (38 20 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
720.77ms 450.61us (38 20 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
721.22ms 247.71us (25 17 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
721.47ms 396.25us (32 21 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
721.87ms 168.99us (20 14 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
722.04ms 1.1160ms (50 38 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
723.16ms 389.30us - - - - - - 2.9898MB 7.4999GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
724.03ms 1.0830ms - - - - - - 5.7678MB 5.2009GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
725.60ms 486.77us - - - - - - 3.1616MB 6.3429GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
727.05ms 1.7477ms - - - - - - 5.6125MB 2.7420GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
729.63ms 163.90us - - - - - - 1.9964MB 11.895GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
731.24ms 5.6033ms - - - - - - 14.624MB 2.5487GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
740.39ms 1.1520us - - - - - - 36B 29.802MB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
740.63ms 704ns - - - - - - 36B 48.767MB/s Pageable Device Tesla T4 (0) 1 7 [CUDA memcopy HtoD]
1.01893s 1.4405ms (20 20 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
1.02037s 2.8228ms (38 20 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
1.02319s 1.8585ms (25 17 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
1.02505s 2.6240ms (32 21 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
1.02768s 980.07us (20 14 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
1.02866s 8.2403ms (50 38 1) (32 32 1) 62 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11EwdRRQPKzLTg
1.03601s 786.96us - - - - - - 2.9898MB 3.7101GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.03918s 1.9321ms - - - - - - 5.7678MB 2.9153GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.04207s 1.5715ms - - - - - - 3.1616MB 1.9646GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.04458s 2.3885ms - - - - - - 5.6125MB 2.1390GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.04701s 162.45us - - - - - - 1.9964MB 11.920GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.04982s 5.6053ms - - - - - - 14.624MB 2.5478GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.05579s 905.00us - - - - - - 2.9898MB 3.2262GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.05816s 2.5379ms - - - - - - 5.7678MB 2.2194GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.06187s 1.5477ms - - - - - - 3.1616MB 1.9949GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.06352s 2.0557ms - - - - - - 5.6125MB 2.3812GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.06601s 162.88us - - - - - - 1.9964MB 11.970GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.06783s 5.4633ms - - - - - - 14.624MB 2.6140GB/s Device Pageable Tesla T4 (0) 1 7 [CUDA memcopy DtoH]
1.34165s 97.981us (20 20 1) (32 32 1) 20 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ewd
1.34177s 190.46us (38 20 1) (32 32 1) 20 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ewd
1.34196s 158.75us (25 17 1) (32 32 1) 20 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ewd
1.34212s 194.68us (32 21 1) (32 32 1) 20 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ewd
1.34232s 66.463us (20 14 1) (32 32 1) 20 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ewd
1.34239s 666.96us (50 38 1) (32 32 1) 20 0B 0B - - - - - Tesla T4 (0) 1 7 _ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ewd
```

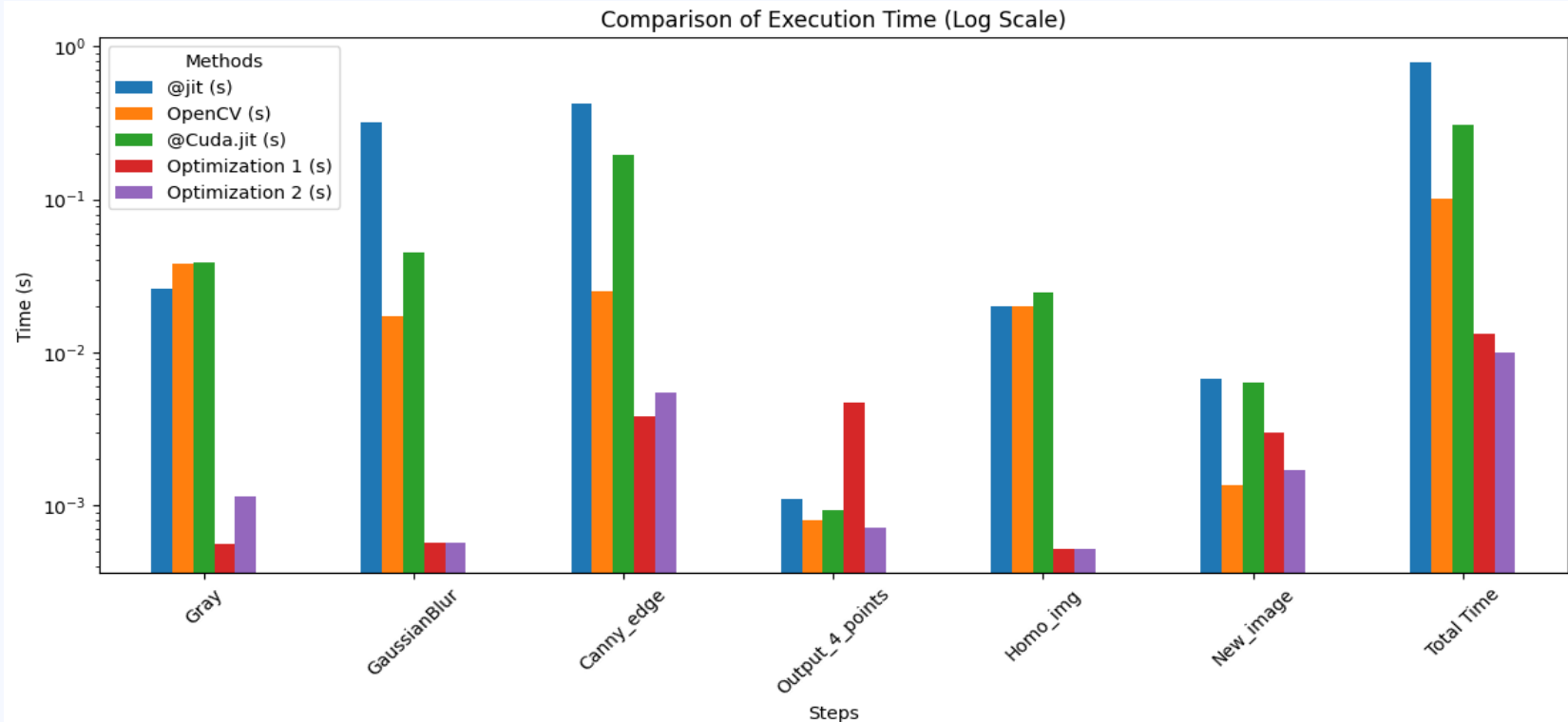
Phiên bản tối ưu lần 2

- Sử dụng CMEM trong CUDA giúp tối ưu hóa hiệu suất của các ứng dụng GPU bằng cách giảm độ trễ truy cập bộ nhớ và tăng băng thông khi dữ liệu được truy cập đồng bộ bởi nhiều luồng.

Phiên bản tối ưu lần 2

```
@cuda.jit
def gradient_kernel_cmem(img_in, img_out, angle, kernel_x, kernel_y):
    c,r = cuda.grid(2)
    kernel_x_cmem = cuda.const.array_like(kernel_x)
    kernel_y_cmem = cuda.const.array_like(kernel_y)
    if r < img_in.shape[0] and c < img_in.shape[1]:
        result_x = 0
        result_y = 0
        for i in range(kernel_x_cmem.shape[0]):
            for j in range(kernel_x_cmem.shape[1]):
                in_r = r - kernel_x_cmem.shape[0]//2 + i
                in_c = c - kernel_x_cmem.shape[1]//2 + j
                in_r = min(max(0, in_r), img_in.shape[0] - 1)
                in_c = min(max(0, in_c), img_in.shape[1] - 1)
                result_x += kernel_x_cmem[i,j] * img_in[in_r,in_c]
                result_y += kernel_y_cmem[i,j] * img_in[in_r,in_c]
        img_out[r,c] = ((result_x ** 2) + (result_y ** 2)) ** 0.5
        angle[r,c] = math.atan2(result_y, result_x) * (180 / np.pi)
        angle[r,c] = angle[r,c] if angle[r,c] >= 0 else angle[r,c]+180
```

Phiên bản tối ưu lần 2



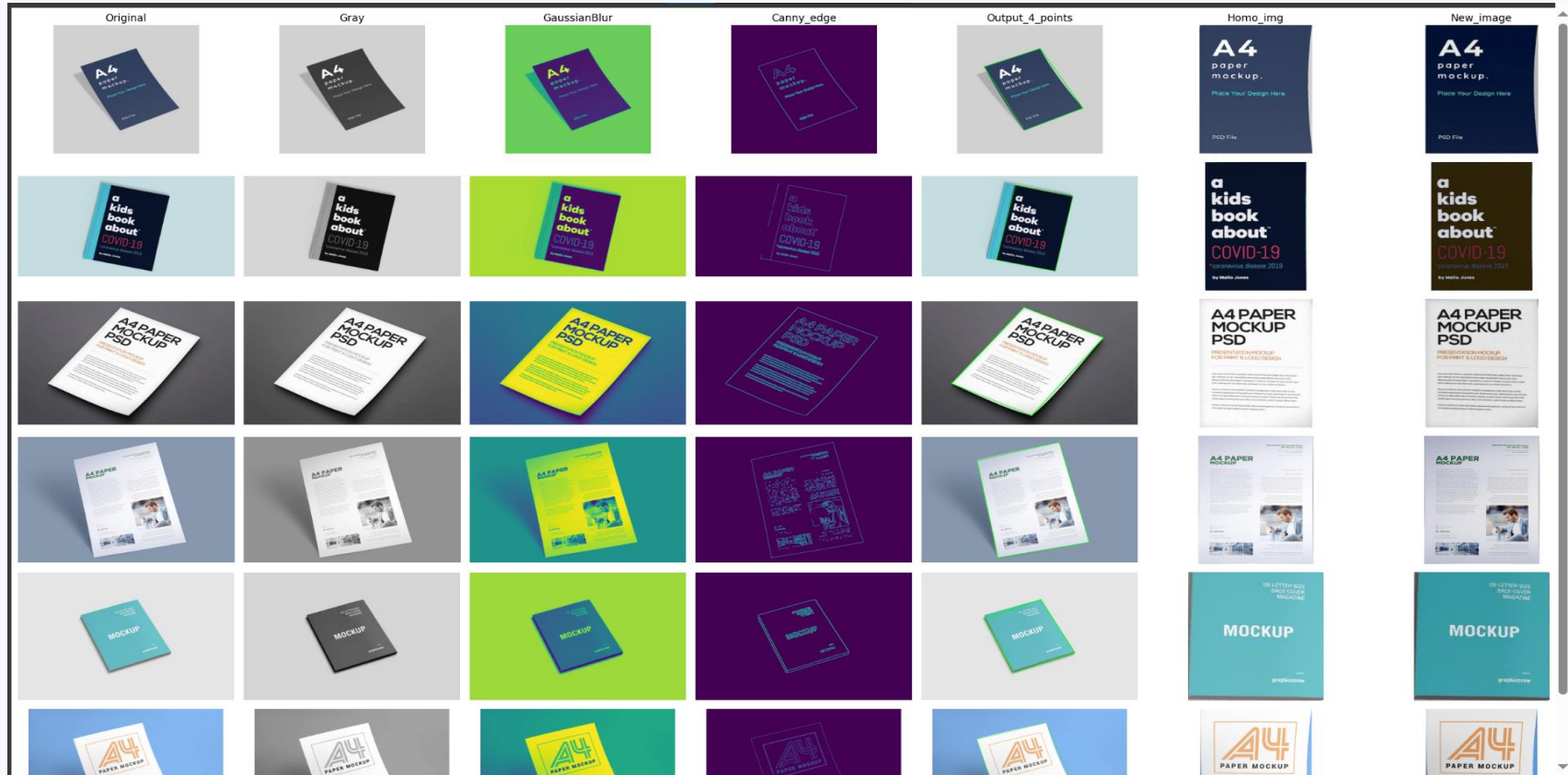
Phiên bản tối ưu lần 2

	@jit (s)	OpenCV (s)	@Cuda.jit (s)	Optimization 1 (s)	Optimization 2 (s)
Gray	0.020392	0.006911	0.053736	0.000551	0.000881
GaussianBlur	0.122405	0.005718	0.046141	0.001040	0.000577
Canny_edge	0.572253	0.011081	0.228040	0.002144	0.002813
Output_4_points	0.000954	0.001349	0.003944	0.000795	0.000700
Homo_img	0.035583	0.034992	0.027035	0.000863	0.000810
New_image	0.005895	0.005548	0.006176	0.001513	0.001412
Total Time	0.757482	0.065599	0.365071	0.006907	0.007195

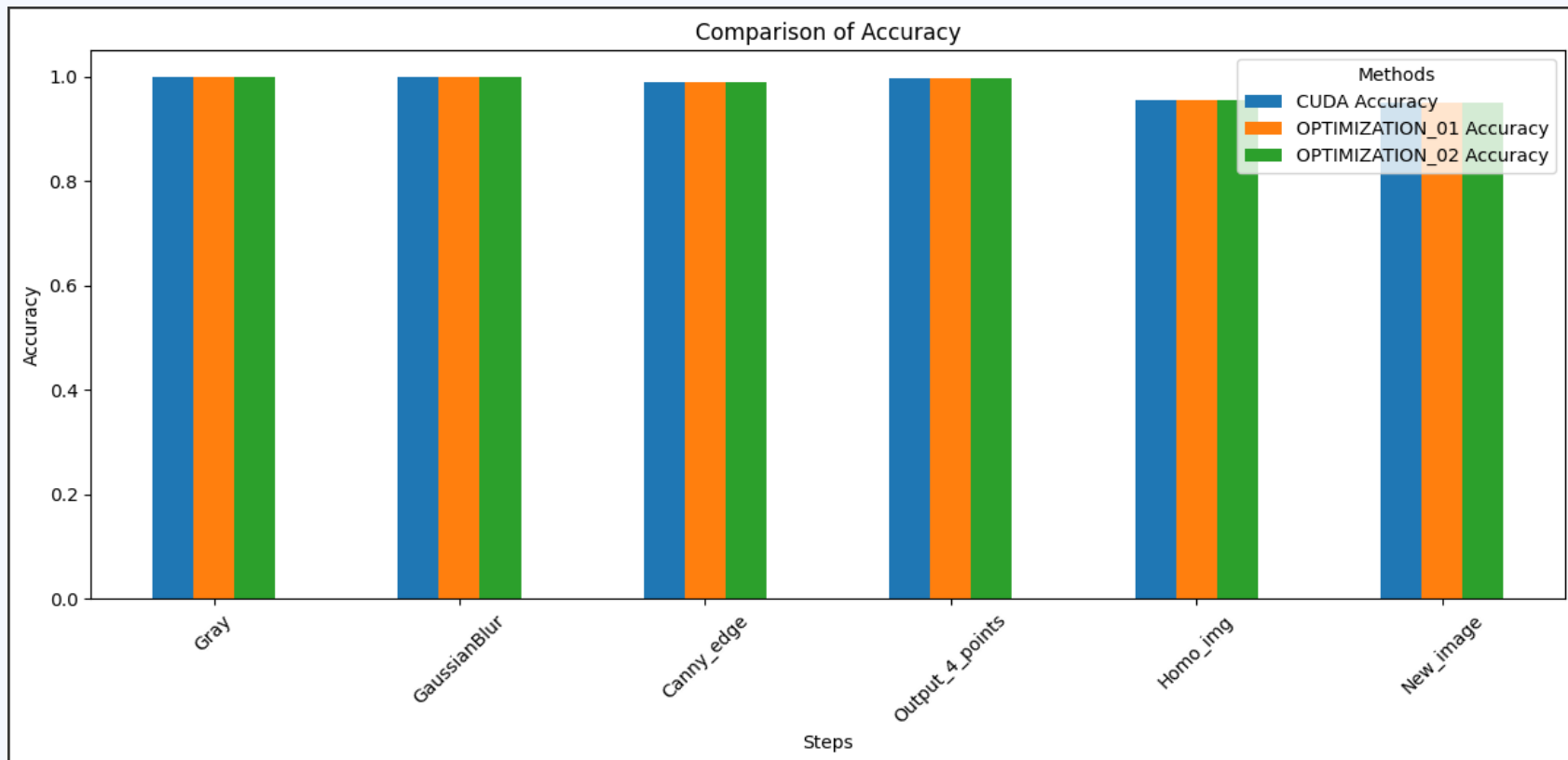
Phiên bản tối ưu lần 2

- Optimization 2 có thời gian thực thi rất nhanh, thường chỉ xấp xỉ 1% đến 5% thời gian của các phương pháp khác như @jit, OpenCV và @Cuda.jit.
- Khi so sánh với Optimization 1, thời gian thực thi của Optimization 2 thường tương đương hoặc chỉ chậm hơn một chút (thường dưới 1.7 lần).
- Optimization 2 cho thấy hiệu suất rất cao, đặc biệt là đối với các phép tính như GaussianBlur và Canny_edge, khi nhanh hơn hàng chục đến hàng trăm lần so với các phương pháp khác.
- Tổng thời gian thực thi của Optimization 2 là nhỏ nhất, chỉ 0.008439 giây, chứng tỏ đây là một trong những phương pháp tối ưu nhất cho các phép tính được xem xét.

Phiên bản tối ưu lần 2



Phiên bản tối ưu lần 2



04

Phiên bản tối ưu lần 3



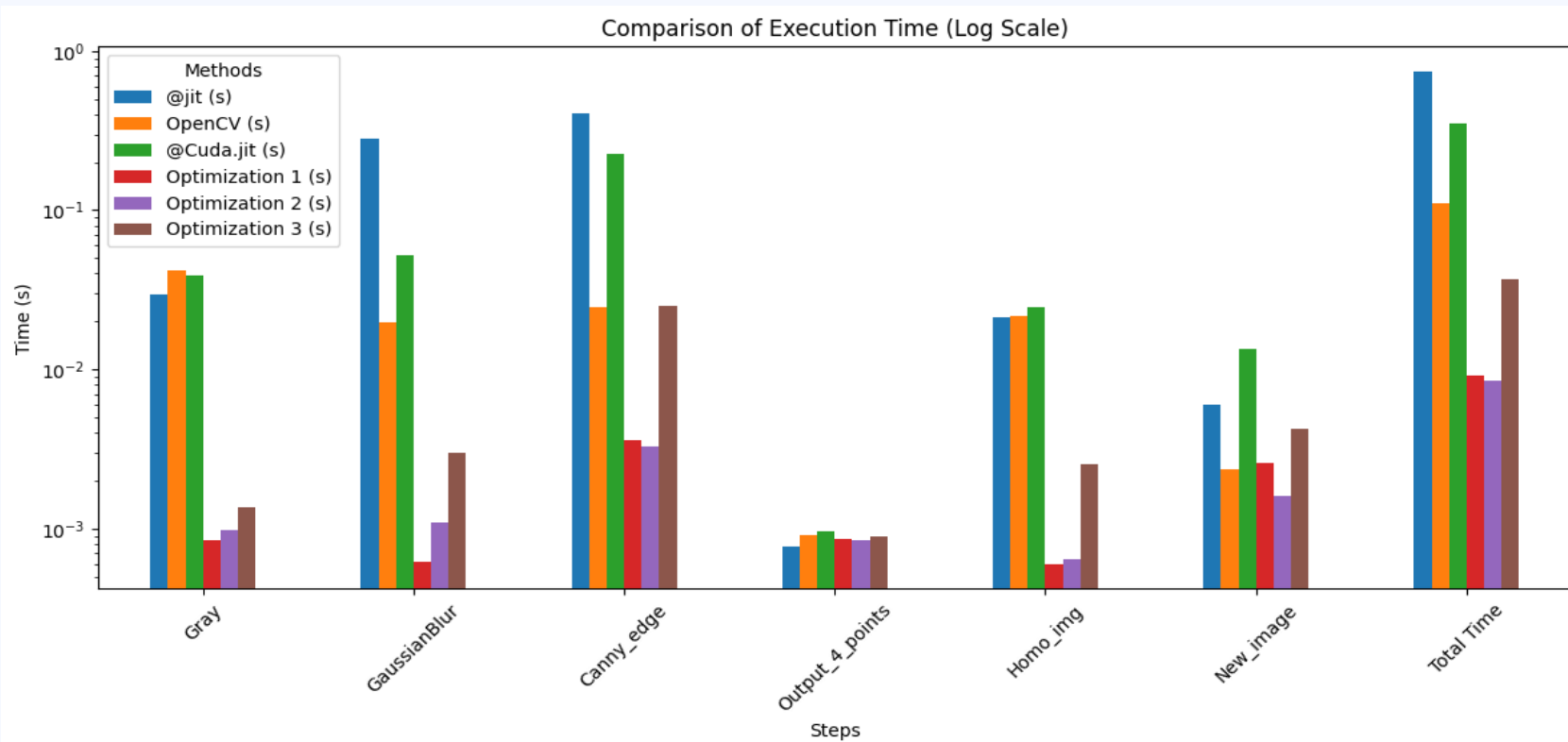
Phiên bản tối ưu lần 3

- Tối ưu hóa bằng cách sử dụng CUDA streams là một phương pháp quan trọng để tăng hiệu suất của các ứng dụng GPU bằng cách thực hiện nhiều tác vụ song song trên GPU.
- CUDA streams cho phép thực hiện chồng chéo giữa tính toán và truyền dữ liệu, cũng như chồng chéo giữa nhiều tác vụ tính toán khác nhau.

Phiên bản tối ưu lần 3

```
d_gray_op3 = [cuda.device_array((d_img[i].shape[0], d_img[i].shape[1]), img[i].dtype) for i in range(len(img))]  
  
grid_size = [(math.ceil(d_gray_op3[i].shape[1] / block_size[0]), #chieu x  
                  math.ceil(d_gray_op3[i].shape[0] / block_size[1])) for i in range(len(img))] #chieu y  
  
streams = [cuda.stream() for i in range(len(img))]  
for i in range(len(img)):  
    convert_RGB_To_Gray_cuda[grid_size[i], block_size, streams[i]](d_img[i], d_gray_op3[i])  
  
    for stream in streams:  
        stream.synchronize()  
  
start_time = time.time()  
for i in range(len(img)):  
    convert_RGB_To_Gray_cuda[grid_size[i], block_size, streams[i]](d_img[i], d_gray_op3[i])  
  
    for stream in streams:  
        stream.synchronize()  
end_time = time.time()  
  
execution_gray_op3_time = end_time - start_time  
  
gray_op3 = [d_gray_op3[i].copy_to_host() for i in range(len(img))]
```

Phiên bản tối ưu lần 3



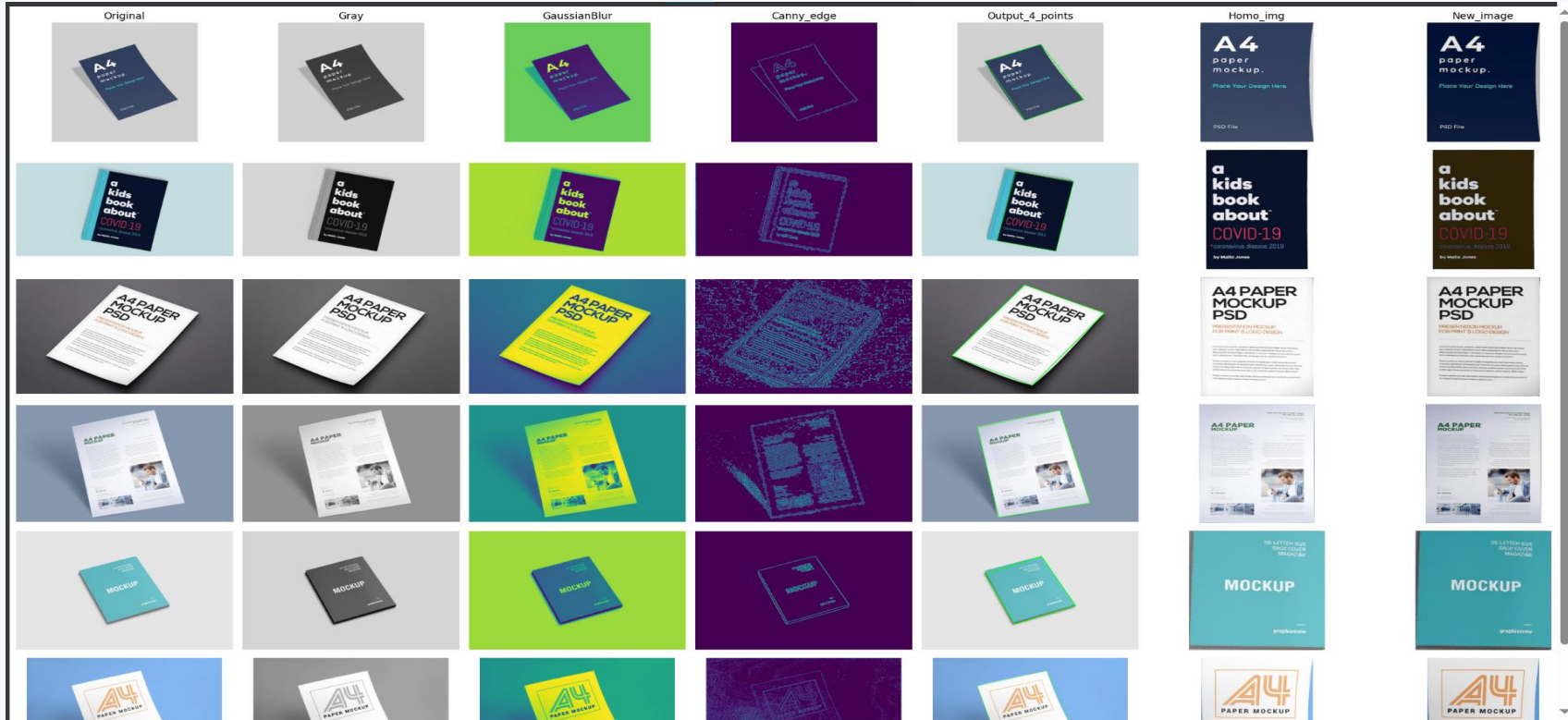
Phiên bản tối ưu lần 3

	@jit (s)	OpenCV (s)	@Cuda.jit (s)	Optimization 1 (s)	Optimization 2 (s)	Optimization 3 (s)
Gray	0.020392	0.006911	0.053736	0.000551	0.000881	0.002318
GaussianBlur	0.122405	0.005718	0.046141	0.001040	0.000577	0.002963
Canny_edge	0.572253	0.011081	0.228040	0.002144	0.002813	0.020866
Output_4_points	0.000954	0.001349	0.003944	0.000795	0.000700	0.000809
Homo_img	0.035583	0.034992	0.027035	0.000863	0.000810	0.002339
New_image	0.005895	0.005548	0.006176	0.001513	0.001412	0.002996
Total Time	0.757482	0.065599	0.365071	0.006907	0.007195	0.032291

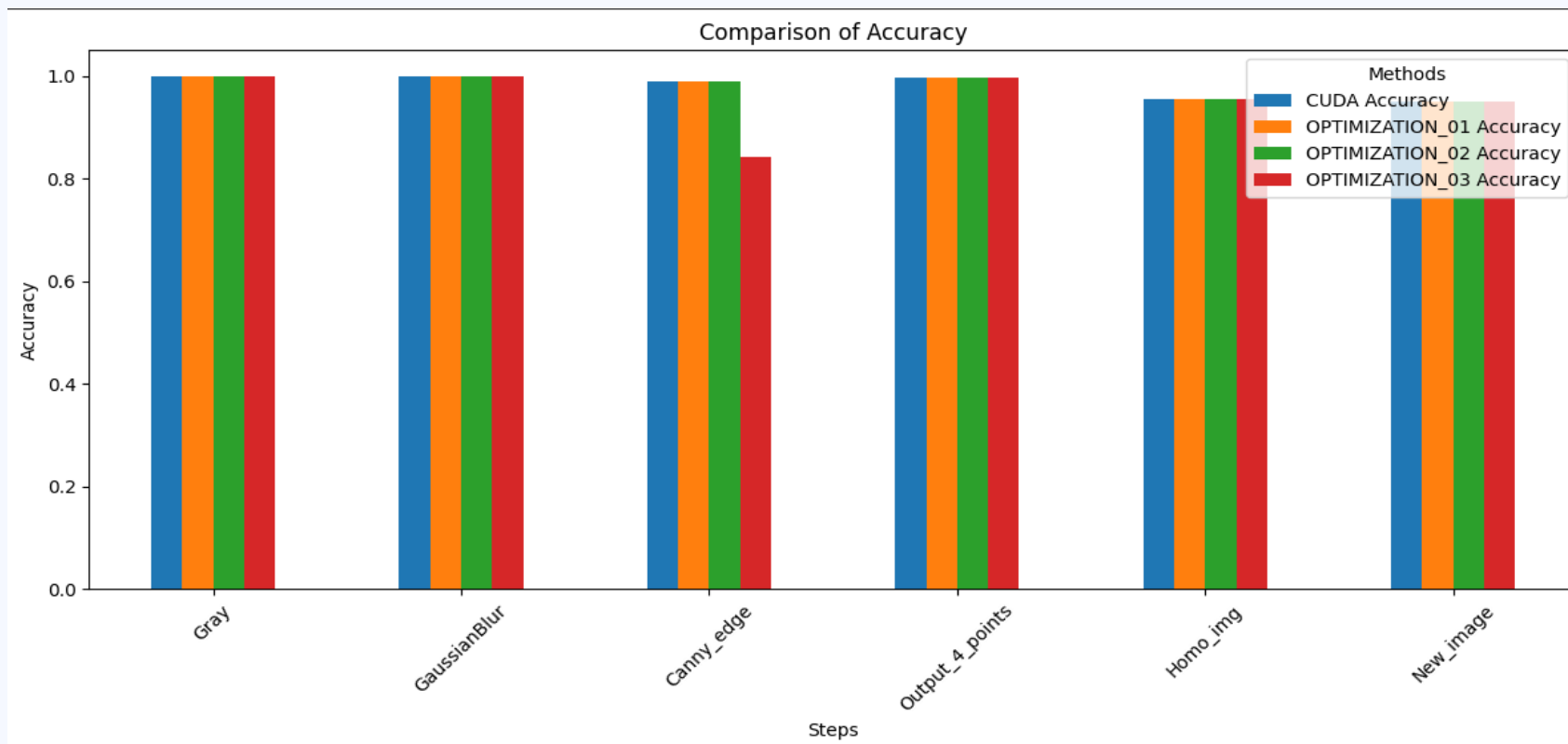
Phiên bản tối ưu lần 3

- Tổng thời gian: Optimization 3 không phải là phương pháp tối ưu nhất, nhanh hơn @jit khoảng 20.1 lần OpenCV khoảng 3 lần và @Cuda.jit khoảng 9.5 lần. Tuy nhiên, chậm hơn Optimization 1 khoảng 4.1 lần và Optimization 2 khoảng 4.4 lần.
- Nhìn chung, Optimization 3 có hiệu suất tốt hơn các phương pháp truyền thống (@jit, OpenCV, @Cuda.jit), nhưng không bằng các phương pháp tối ưu hóa khác (Optimization 1 và Optimization 2).

Phiên bản tối ưu lần 3



Phiên bản tối ưu lần 3

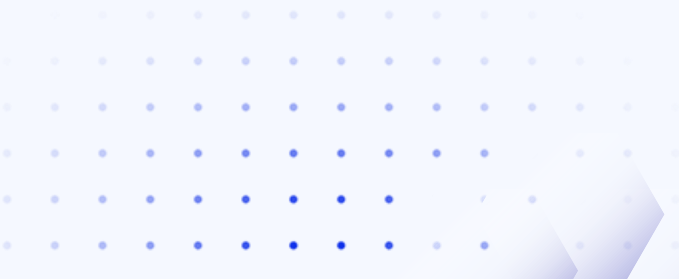
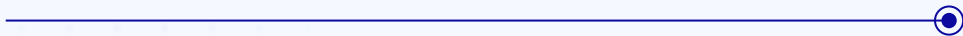


Phiên bản tối ưu lần 3

--1648--	NVPProf is profiling process 1648, command: python3 parallel_optimize_03.py --config config.json														
--1648--	Profiling application: python3 parallel_optimize_03.py --config config.json														
--1648--	Profiling result:														
Start	Duration	Grid Size	Block Size	Regs*	SSMem*	DSMem*	Size	Throughput	SrcMemType	DstMemType	Device	Context	Stream	Name	
237.07ms	124.32us	-	-	-	-	-	1.1212MB	8.8072GB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
238.86ms	317.91us	-	-	-	-	-	2.1629MB	6.6441GB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
240.01ms	120.48us	-	-	-	-	-	1.1856MB	9.6183GB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
242.00ms	274.43us	-	-	-	-	-	1.8797MB	6.6890GB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
242.81ms	95.214us	-	-	-	-	-	766.61KB	8.5795GB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
245.41ms	1.9015ms	-	-	-	-	-	5.4840MB	5.3474GB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
505.07ms	88.381us	(20 20 1)	(32 32 1)	17	0B	0B	-	-	-	-	Tesla T4 (0)	1	13	ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ew	
505.20ms	165.34us	(38 20 1)	(32 32 1)	17	0B	0B	-	-	-	-	Tesla T4 (0)	1	14	ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ew	
505.36ms	92.286us	(25 17 1)	(32 32 1)	17	0B	0B	-	-	-	-	Tesla T4 (0)	1	15	ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ew	
505.45ms	147.29us	(32 21 1)	(32 32 1)	17	0B	0B	-	-	-	-	Tesla T4 (0)	1	16	ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ew	
505.60ms	65.566us	(20 14 1)	(32 32 1)	17	0B	0B	-	-	-	-	Tesla T4 (0)	1	17	ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ew	
505.65ms	412.76us	(50 38 1)	(32 32 1)	17	0B	0B	-	-	-	-	Tesla T4 (0)	1	18	ZN6cudapy8_main_24convert_RGB_To_Gray_cuda82v1B94cw51cXTLSUwv1sCUT9Uw11Ew	
506.10ms	88.830us	-	-	-	-	-	-	-	-	-	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
506.45ms	166.17us	-	-	-	-	-	382.60KB	4.1086GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
506.88ms	98.285us	-	-	-	-	-	738.28KB	4.2371GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
507.12ms	151.90us	-	-	-	-	-	404.69KB	3.9299GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
507.50ms	55.263us	-	-	-	-	-	641.60KB	4.0282GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
507.70ms	456.76us	-	-	-	-	-	255.54KB	4.4098GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
512.02ms	1.2480us	-	-	-	-	-	1.8280MB	3.9083GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
660.13ms	230.93us	(20 20 1)	(32 32 1)	62	0B	0B	72B	55.02MB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
660.35ms	455.89us	(38 20 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	19	ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz	
660.80ms	252.76us	(25 17 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	20	ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz	
661.04ms	397.08us	(32 21 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	21	ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz	
661.43ms	178.56us	(20 14 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	22	ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz	
661.59ms	1.1218ms	(50 38 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	23	ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz	
662.83ms	1.2084ms	-	-	-	-	-	-	-	-	-	Tesla T4 (0)	1	24	ZN6cudapy8_main_16convolution_cuda82v2B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz	
665.43ms	2.2215ms	-	-	-	-	-	2.9898MB	2.4161GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
669.70ms	1.8119ms	-	-	-	-	-	5.7678MB	1.9817GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
671.64ms	2.5647ms	-	-	-	-	-	3.1615MB	1.6947GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
675.09ms	486.49us	-	-	-	-	-	5.0125MB	1.9086GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
677.08ms	6.4222ms	-	-	-	-	-	1.9964MB	4.0075GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
686.95ms	1.2160us	-	-	-	-	-	14.624MB	2.2238GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
687.21ms	704ns	-	-	-	-	-	36B	28.234MB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
976.91ms	1.4414ms	(20 20 1)	(32 32 1)	62	0B	0B	36B	48.767MB/s	Pageable	Device	Tesla T4 (0)	1	7	[CUDA memcopy HtoD]	
979.20ms	2.8617ms	(38 20 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	25	ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz1Tg	
981.06ms	1.9120ms	(25 17 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	26	ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz1Tg	
982.92ms	2.6500ms	(32 21 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	27	ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz1Tg	
985.50ms	1.0065ms	(20 14 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	28	ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz1Tg	
986.41ms	8.2849ms	(50 38 1)	(32 32 1)	62	0B	0B	-	-	-	-	Tesla T4 (0)	1	30	ZN6cudapy8_main_13gradient_cuda82v3B94cw51cXTLSUwv1sCUT9Uw11Ew1dRRQPkz1Tg	
994.85ms	1.1843ms	-	-	-	-	-	2.9898MB	2.4654GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
997.43ms	2.8423ms	-	-	-	-	-	5.7678MB	1.9817GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.00166s	1.8210ms	-	-	-	-	-	3.1615MB	1.6947GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.00448s	2.6066ms	-	-	-	-	-	5.0125MB	1.8779GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.00746s	475.32us	-	-	-	-	-	1.9964MB	4.0166GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.00937s	6.4104ms	-	-	-	-	-	14.624MB	2.2278GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.01685s	1.3371ms	-	-	-	-	-	2.9898MB	2.1836GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.01964s	2.6280ms	-	-	-	-	-	5.7678MB	2.1433GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.02299s	2.0266ms	-	-	-	-	-	3.1616MB	1.5235GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.02624s	2.2263ms	-	-	-	-	-	5.0125MB	2.1088GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.02956s	475.57us	-	-	-	-	-	1.9964MB	4.0994GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.03169s	6.3495ms	-	-	-	-	-	14.624MB	2.2492GB/s	Device	Pageable	Tesla T4 (0)	1	7	[CUDA memcopy DtoH]	
1.32700s	98.239us	(20 20 1)	(32 32 1)	20	0B	0B	-	-	-	-	Tesla T4 (0)	1	31	ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ew1d	
1.32722s	190.84us	(38 20 1)	(32 32 1)	20	0B	0B	-	-	-	-	Tesla T4 (0)	1	32	ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ew1d	
1.32741s	160.25us	(25 17 1)	(32 32 1)	20	0B	0B	-	-	-	-	Tesla T4 (0)	1	33	ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ew1d	
1.32756s	195.16us	(32 21 1)	(32 32 1)	20	0B	0B	-	-	-	-	Tesla T4 (0)	1	34	ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ew1d	
1.32775s	66.942us	(20 14 1)	(32 32 1)	20	0B	0B	-	-	-	-	Tesla T4 (0)	1	35	ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ew1d	
1.32798s	665.97us	(50 38 1)	(32 32 1)	20	0B	0B	-	-	-	-	Tesla T4 (0)	1	36	ZN6cudapy8_main_22nonMaxSuppression_cuda82v5B94cw51cXTLSUwv1sCUT9Uw11Ew1d	

05

Tổng kết



Tổng kết

- Trong 4 phiên bản song song (Cơ bản, tối ưu 1, tối ưu 2, tối ưu 3) thì phiên bản song song tối ưu 1 mang lại kết quả tốt nhất
- Nguyên nhân hiệu suất của các phiên bản tối ưu 3 không tốt có thể là do việc sử dụng mảng NumPy phải đi qua các API trừu tượng hóa cũng như việc cho phép chồng chéo (overlap) giữa việc truyền dữ liệu và việc thực thi kernel còn hạn chế.
- Việc giảm thiểu copy dữ liệu CPU-GPU và tối ưu truy cập bộ nhớ là những yếu tố quan trọng để cải thiện hiệu suất của ứng dụng CUDA trong python.

**Cảm ơn thầy và các
bạn đã lắng nghe !!!**

