	Result	Size	Time	Cycles	GPU	SM Frequency	Process	Attributes
Current	576 - convolveHorizontalKernel	(20, 15, 1)x(16, 16, 1)	6.85 us	5,285	0 - NVIDIA GeForce RTX 3070 Laptop GPU	771.59 Mhz	[2924] example2	

Summary | Details | Source | Context | Comments | Raw | Session

Compare | Tools | View | Export

This table shows all results in the report. Use the column headers to sort the results in this report. Double-click a result to see detailed metrics. Double-click on demangled names to rename it.

ID	Estimated Speedup [%]	Function Name	Demangled Name	Duration [us] (319.74 us)	Runtime Improvement [us] (159.87 us)	Compute Throughput [%]	Memory Throughput [%]	# Reg
0	50.00	**convolveHorizo...**	**convolveHorizo...**	6.85	3.42	33.53	33.53	
1	50.00	convolveVerticalKe...	convolveVerticalKe...	6.98	3.49	34.14	34.14	
2	50.00	convolveHorizonta...	convolveHorizonta...	7.68	3.84	38.25	38.25	
3	50.00	convolveVerticalKe...	convolveVerticalKe...	8.74	4.37	34.45	34.45	
4	50.00	convolveHorizonta...	convolveHorizonta...	7.74	3.87	38.03	38.03	
5	50.00	convolveVerticalKe...	convolveVerticalKe...	8.58	4.29	34.19	34.19	
6	50.00	featureStrengthCo...	featureStrengthCo...	21.12	10.56	55.36	57.30	
7	50.00	convolveHorizonta...	convolveHorizonta...	6.85	3.42	34.59	34.59	
8	50.00	convolveVerticalKe...	convolveVerticalKe...	6.75	3.38	33.40	33.40	
9	50.00	convolveHorizonta...	convolveHorizonta...	11.97	5.98	61.69	61.69	
10	50.00	convolveVerticalKe...	convolveVerticalKe...	13.09	6.54	52.85	52.85	
11	50.00	convolveHorizonta...	convolveHorizonta...	7.71	3.86	38.27	38.27	
12	50.00	convolveVerticalKe...	convolveVerticalKe...	8.74	4.37	34.07	34.07	

The following performance optimization opportunities were discovered for this result. Follow the rule links to see more context on the Details page.
Note: *Speedup estimates provide upper bounds for the optimization potential of a kernel assuming its overall algorithmic structure is kept unchanged.*

Tail Effect
Est. Speedup: 50.00%

A wave of thread blocks is defined as the maximum number of blocks that can be executed in parallel on the target GPU. The number of blocks in a wave depends on the number of multiprocessors and the theoretical occupancy of the kernel. This kernel launch results in 1 full waves and a partial wave of 60 thread blocks. Under the assumption of a uniform execution duration of all thread blocks, this partial wave may account for up to 50.0% of the total runtime of this kernel. Try launching a grid with no partial wave. The overall impact of this tail effect also lessens with the number of full waves executed for a grid. See the Hardware Model description for more details on launch configurations.

▸ Key Performance Indicators