	Result	Size	Time	Cycles	GPU	SM Frequency	Process	Attributes
Current	576 - convolveHorizontalKernel	(20, 15, 1)x(16, 16, 1)	6.78 us	5,245	0 - NVIDIA GeForce RTX 3070 Laptop GPU	772.75 Mhz	[3339] example3	

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] (994.08 us)	Runtime Improvement [us] (497.04 us)	Compute Throughput [%]	Memory Throughput [%]	# Reg
0	50.00	**convolveHorizo...**	**convolveHorizo...**	6.78	3.39	34.45	34.45	
1	50.00	convolveVerticalKe...	convolveVerticalKe...	6.91	3.46	33.34	33.34	
2	50.00	convolveHorizonta...	convolveHorizonta...	7.94	3.97	38.96	38.96	
3	50.00	convolveVerticalKe...	convolveVerticalKe...	8.83	4.42	32.91	32.91	
4	50.00	convolveHorizonta...	convolveHorizonta...	7.74	3.87	38.96	38.96	
5	50.00	convolveVerticalKe...	convolveVerticalKe...	8.77	4.38	33.24	33.24	
6	50.00	featureStrengthCo...	featureStrengthCo...	20.77	10.38	53.54	55.44	
7	50.00	convolveHorizonta...	convolveHorizonta...	6.88	3.44	33.94	33.94	
8	50.00	convolveVerticalKe...	convolveVerticalKe...	6.91	3.46	33.51	33.51	
9	50.00	convolveHorizonta...	convolveHorizonta...	11.90	5.95	61.68	61.68	
10	50.00	convolveVerticalKe...	convolveVerticalKe...	12.86	6.43	54.24	54.24	
11	50.00	convolveHorizonta...	convolveHorizonta...	7.71	3.86	38.93	38.93	
12	50.00	convolveVerticalKe...	convolveVerticalKe...	8.74	4.37	33.81	33.81	

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

L2 Slices Workload Imbalance One or more L2 Slices have a much higher number of active cycles than the average number of active cycles. Maximum instance value is 14.17% above the average, while