

week-3

JX-Ma

2024/3/16

1 本周工作

1. 更新了一下数据布局图，有直接卷积，im2win,im2col 还有目前正在使用的数据布局。
2. 重新跑了一遍上周的实验和直接卷积，记录了他们最大的 gflops。
3. 整理了一下直接卷积和 im2col 的内存对比
4. 查看高性能计算平台 cpu 信息

2 数据布局

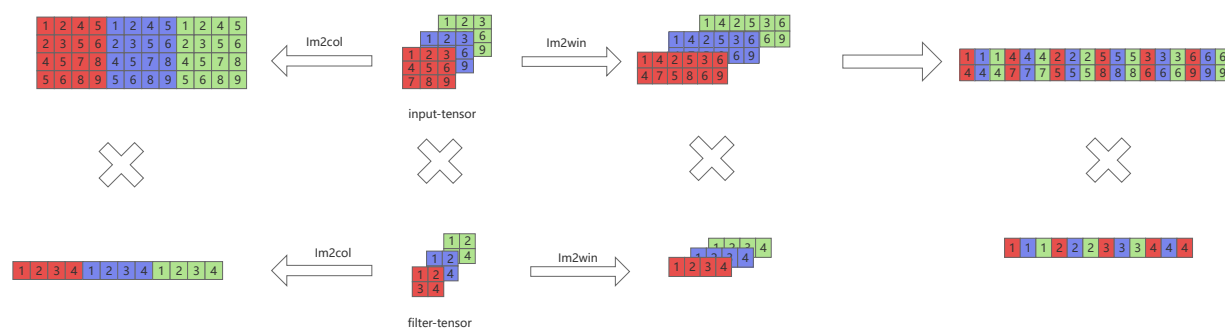


图 1: layout

我和张新鹏使用的布局差不多，唯一一个区别的是按照输入张量通道展开的大小不同，我做的实验室，

对于前面几个 benchmark 它的输入张量的 channel 不能被 8 整除的就没有展开，而后面几个 benchmark 中的 channel 能被 8 整除的则按照 8 展开。张新鹏做的是全部按照 channel 展开。

3 实验部分

3.1 实验环境

- 系统: Ubuntu 22.01
- gcc version : 9.5.0
- 优化选项: -O3 -fopenmp -avx2 -fmadd
- cpu:AMD Ryzen 7 6800H 3.20GHz

3.2 实验结果

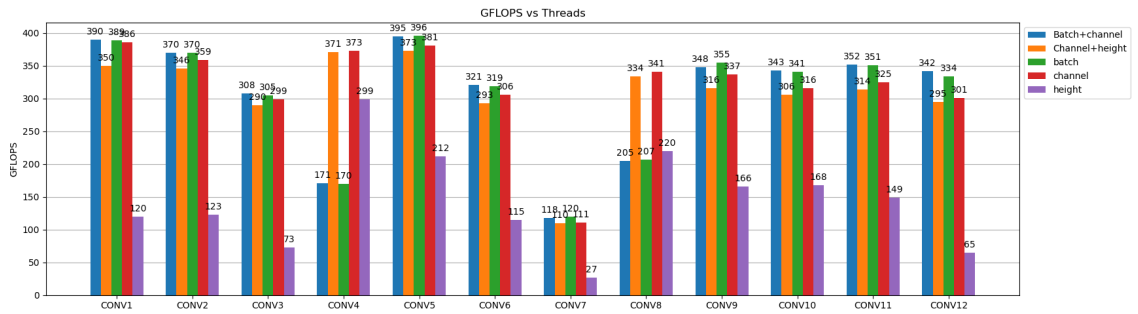


图 2: directConvolution-gflops

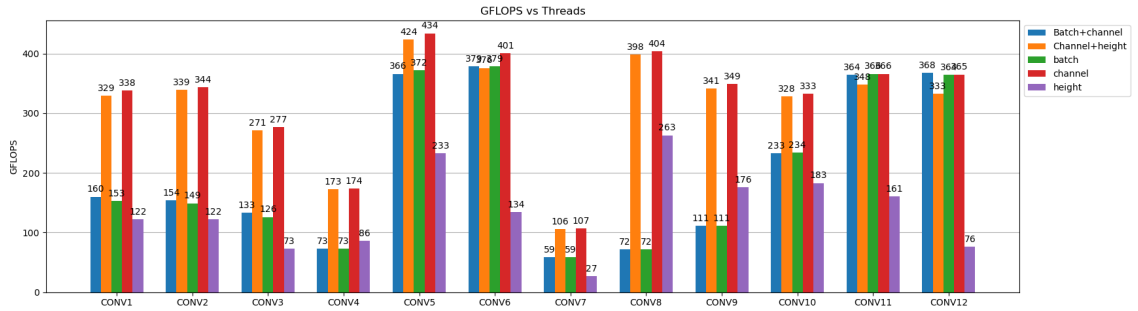


图 3: im2winConvolution-gflops

3.3 实验分析

从结果上看，在直接卷积上 openmp 在 batch, channel, batch+channel 上效果差不多，im2win 上在 channel 这一层循环外加上 openmp 效果最好。

4 内存分析

- param1:Percent of CPU this job got
- param2:Maximum resident set size (kbytes)
- param3:Major (requiring I/O) page faults
- param4:Minor (reclaiming a frame) page faults
- param5:Voluntary context switches
- param6:Involuntary context switches

指标	含义
Percent of CPU this job got	命令执行时 CPU 的占用比例
Maximum resident set size	执行程序所占用内存的最大值。
Major (requiring I/O) page faults	此程序的主要内存页错误发生的次数。
Minor (reclaiming a frame) page faults	此程序的次要内存页错误发生的次数。
Voluntary context switches	此程序自愿中断（I/O 执行完毕，磁碟读取完成等）的次数
Involuntary context switches	此程序被强迫中断（如 CPU 时间耗尽）的次数

表 1: CONV1-memory-analyze

补充：

1. Percent of CPU this job got 是内核模式的 CPU 时间加上使用者模式的 CPU 时间除以总时间
 2. swap 分区：Swap space 交换空间，是虚拟内存的表现形式。系统为了应付一些需要大量内存的应用，而将磁盘上的空间做内存使用，当物理内存不够用时，将其中一些暂时不需要的数据交换到交换空间，也叫交换文件或页面文件中。

3. Major (requiring I/O) page faults: 是指某一内存页已经置换到 SWAP 分区中，又被其他程序使用过，该页的内容必须从 SWAP 分区里再读出来才能使用

4. Minor (reclaiming a frame) page faults: 是指某一内存页虽然已经置换到 SWAP 中，但尚未被其他程序使用。此时该页的内容并未被破坏，不必从 SWAP 分区里读出来即可直接使用

如下图，我的理解内存和 swap 之间有一个 cache，它会把最近使用到的页面从 swap 交换区中拿到，然后内存先读取 cache，没读取到才会去访问 swap 交换区。主要页面错误说明需要的数据在 swap 交换区内，并且在 cache 中，但是该数据已经被其他进程使用过了，数据可能会被修改，于是需要从 swap 中获取干净的数据。次要页面错误则是需要的数据在 cache 中且没有被其他进程应用，所以可以直接从 cache 中获取。

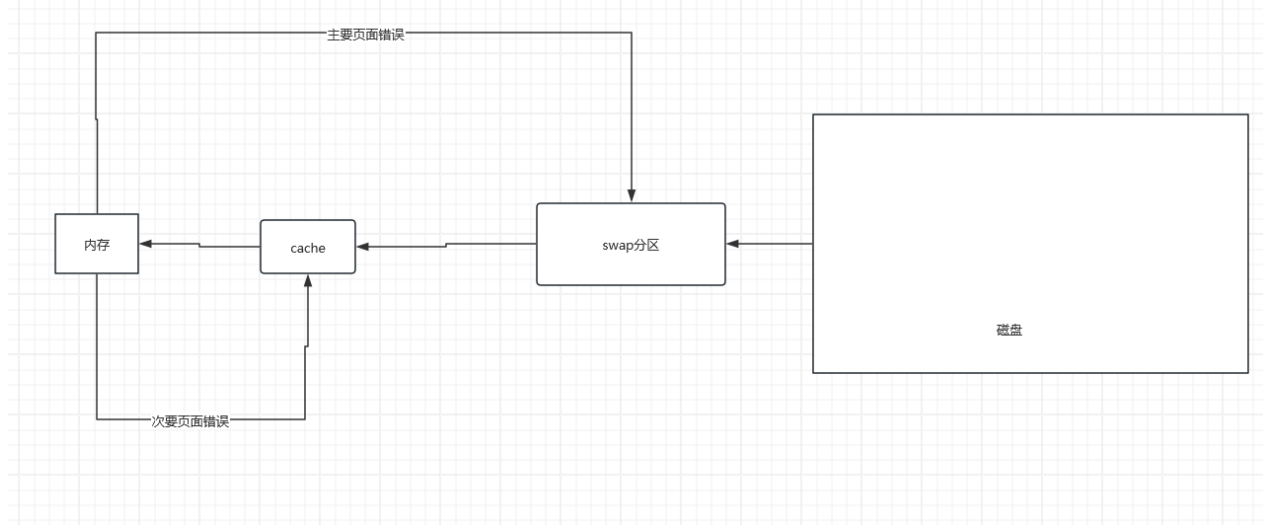


图 4: major page Fault & minor page Fault

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	706	309900	0	87763	34	135	389	direct
	841	438664	3	119947	40	1212	153	im2win
Channel	717	310152	0	87847	39	174	386	direct
	722	438536	4	120029	45	126	338	im2win
Height	1062	310024	1	467675	180	916	120	direct
	1208	438920	0	491857	695	3150	122	im2win
Batch+Channel	716	309900	0	87825	53	167	390	direct
	481	438668	10	119955	80	116	160	im2win
channel+height	732	310408	0	87947	53	104	350	direct
	768	439304	3	120127	68	457	329	im2win

表 2: CONV1-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	737	317832	0	90468	42	225	370	direct
	912	451592	0	123822	45	383	149	im2win
Channel	739	318472	1	90594	36	530	359	direct
	743	451720	1	123945	39	339	344	im2win
Height	975	317960	4	470230	40	395	123	direct
	1243	451336	0	504814	442	3376	122	im2win
Batch+Channel	716	317960	0	90535	96	667	370	direct
	900	451724	0	123867	62	400	154	im2win
channel+height	774	318604	0	90684	56	468	346	direct
	721	452104	1	124069	54	555	339	im2win

表 3: CONV2-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	497	957320	0	249313	53	497	305	direct
	1292	17051044	0	4562748	44	5455	126	im2win
Channel	509	956808	0	249370	40	250	299	direct
	1168	17050828	10	4563471	136	3285	277	im2win
Height	886	956680	0	1024213	76	453	73	direct
	1403	17051144	0	16976644	505	9759	73	im2win
Batch+Channel	505	957320	0	249371	58	359	308	direct
	1279	17051660	0	4562855	64	6122	133	im2win
channel+height	526	957320	0	249474	56	248	290	direct
	1147	17051528	0	4563486	124	4167	271	im2win

表 4: CONV3-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	1153	2464776	11	1008548	65	3992	170	direct
	1471	6328584	0	1974460	73	11273	73	im2win
Channel	958	2464520	0	1008545	52	1980	373	direct
	1390	6328456	0	1974460	238	2763	174	im2win
Height	1058	2464776	0	1316390	144	6220	299	direct
	1478	6328584	0	2260323	466	8914	86	im2win
Batch+Channel	1166	2465288	0	1008608	69	3247	171	direct
	1453	6328712	0	1974521	80	10355	73	im2win
channel+height	965	2465292	0	1008652	58	2317	371	direct
	1385	6328840	0	1974572	354	4729	173	im2win

表 5: CONV4-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	839	168332	1	40453	50	382	396	direct
	797	256648	0	62337	49	301	372	im2win
Channel	917	168584	0	49236	35	164	381	direct
	853	256140	0	72693	121	177	434	im2win
Height	1102	168840	0	119062	77	687	212	direct
	1209	256264	4	144832	611	1576	233	im2win
Batch+Channel	877	168584	0	40511	54	334	395	direct
	790	256652	0	62393	95	1327	366	im2win
channel+height	946	169352	0	49832	75	282	373	direct
	863	256776	8	73086	165	711	424	im2win

表 6: CONV5-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	751	136072	0	30545	46	955	319	direct
	703	163720	0	37459	36	194	379	im2win
Channel	801	136076	0	37591	39	128	306	direct
	857	163976	0	47949	130	492	401	im2win
Height	1030	136072	0	59962	36	549	115	direct
	1212	163848	3	66894	527	2517	134	im2win
Batch+Channel	803	135820	0	30612	61	131	321	direct
	667	164236	6	37519	60	896	379	im2win
channel+height	814	136712	0	37721	58	128	293	direct
	709	164232	0	48083	58	262	376	im2win

表 7: CONV6-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	383	3331852	0	842352	52	344	120	direct
	954	3479944	0	879478	51	729	59	im2win
Channel	408	3331080	0	842464	37	176	111	direct
	917	3479304	0	879590	142	698	107	im2win
Height	840	3331212	0	2289075	176	2681	27	direct
	1267	3479560	0	2326743	514	4594	27	im2win
Batch+Channel	400	3332108	0	842412	72	182	118	direct
	981	3480208	1	879532	73	1179	379	im2win
channel+height	410	3331336	0	842567	55	169	59	direct
	924	3480072	0	879659	120	706	106	im2win

表 8: CONV7-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	1200	1263624	4	407193	62	7766	207	direct
	1236	2045448	0	602395	49	4381	72	im2win
Channel	1193	1263628	0	407248	37	463	341	direct
	763	2044940	0	602465	38	1081	404	im2win
Height	1302	1263496	0	1320448	89	1120	220	direct
	1038	2045064	0	1610555	211	4962	263	im2win
Batch+Channel	1265	1264008	0	407250	72	2017	205	direct
	1213	2045328	0	602454	90	6284	72	im2win
channel+height	1210	1263880	0	407341	69	503	334	direct
	773	2045448	0	602518	55	2204	398	im2win

表 9: CONV8-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	699	280972	2	86256	44	1119	355	direct
	782	471304	0	133681	43	456	111	im2win
Channel	765	281352	2	86288	41	87	337	direct
	706	471560	0	133708	166	675	349	im2win
Height	935	281224	0	281397	63	1236	166	direct
	1091	471304	1	343795	531	2235	176	im2win
Batch+Channel	761	280968	0	86312	51	112	348	direct
	727	471180	3	133746	65	1442	111	im2win
channel+height	734	281992	0	86380	55	95	316	direct
	671	471688	0	133817	135	320	341	im2win

表 10: CONV9-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	721	180876	1	48852	52	536	341	direct
	677	270988	12	71233	57	320	234	im2win
Channel	752	181512	0	53217	37	124	316	direct
	749	271112	7	75788	94	471	333	im2win
Height	877	181000	1	121777	36	2653	168	direct
	1143	271112	1	148434	451	2466	183	im2win
Batch+Channel	739	182152	0	48895	56	225	343	direct
	690	270984	0	71272	99	568	233	im2win
channel+height	675	181956	0	53211	56	185	306	direct
	755	271500	0	75726	144	343	328	im2win

表 11: CONV10-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	695	133000	0	30931	38	466	351	direct
	625	173708	0	40785	36	175	366	im2win
Channel	724	133384	0	37494	36	131	325	direct
	598	172684	0	47798	40	355	366	im2win
Height	842	133260	0	53958	49	392	149	direct
	1175	172680	12	63772	509	2315	161	im2win
Batch+Channel	716	133516	0	30992	53	98	352	direct
	625	173708	0	40850	64	354	364	im2win
channel+height	714	133768	0	37746	92	736	314	direct
	647	173324	0	48681	55	212	348	im2win

表 12: CONV11-memory-analyze

DIMENTION	param1	param2	param3	param4	param5	param6	gflops	layout
Batch	607	115708	1	25104	37	495	334	direct
	603	130572	12	28697	50	1675	364	im2win
Channel	694	115720	0	31671	36	392	301	direct
	608	130184	0	36683	37	113	365	im2win
Height	1011	115336	0	28304	48	350	65	direct
	1198	129928	0	31889	550	2951	76	im2win
Batch+Channel	691	116104	11	25165	66	98	342	direct
	617	130572	0	28748	57	279	368	im2win
channel+height	711	116232	4	31162	60	416	295	direct
	641	130696	0	36186	76	246	333	im2win

表 13: CONV12-memory-analyze