Convolution

LeNet-5	MNIST				Single batch size						GEMM size						
	Types	Inputs			Weights			Outputs									
		H	W	C	R	S	K	P	Q	K	m (P*Q)	n (K)	k (R*S*C)				
Layer 1	Convolution	32	32	1	5	5	6	28	28	6	784	6	25	235200	97816	2.404515	TRUE
Layer2	subsampling	28	28	6	2	2	6	14	14	6							
Layer3	convolution	14	14	6	5	5	16	10	10	16	100	16	150	480000	76000	6.315789	TRUE
Layer4	subsampling	10	10	16	2	2	16	5	5	16							
Layer5	Convolution	5	5	16	5	5	120	1	1	120	1	120	400	96000	194080	0.494641	TRUE
Layer6	fully connected	1	1	120	1	1	84	1	1	84	1	84	120	20160	41136	0.490082	TRUE
Layer7	fully connected	1	1	84	1	1	10	1	1	10	1	10	84	1680	3736	0.449679	TRUE

AlexNet-5	ImageNet				Single batch size												
Layer 1	Conv	227	227	3	11	11	96	55	55	96	3025	96	363	210830400	5693292	37.03137	FALSE
Layer2	Pool																
Layer3	Conv	27	27	96	5	5	256	27	27	256	729	256	2400	895795200	10202496	87.80157	FALSE
Layer4	Pool																
Layer5	Conv	13	13	256	3	3	384	13	13	384	169	384	2304	299040768	5356032	55.83252	FALSE
Layer6	Conv	13	13	384	3	3	384	13	13	384	169	384	3456	448561152	7904256	56.74932	FALSE
Layer7	Conv	13	13	384	3	3	256	13	13	256	169	256	3456	299040768	6048256	49.44248	FALSE
Layer8	Pool																
Layer9	FC	7	7	512	7	7	4096	1	1	4096	1	4096	25088	205520896	4.11E+08	0.499858	TRUE
Layer10	FC	1	1	4096	1	1	4096	1	1	4096	1	4096	4096	33554432	67141632	0.499756	TRUE
Layer11	FC	1	1	4096	1	1	1000	1	1	1000	1	1000	4096	8192000	16404384	0.499379	TRUE

ResNet-18	ImageNet				Single batch size												
		H	W	C	R	S	K	P	Q	K	m (P*Q)	n (K)	k (R*S*C)				
Layer 1	conv	227	227	3	7	7	64	112	112	64	12544	64	147	236027904	10624768	22.21488	TRUE
Layer 2	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 3	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 4	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 5	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 6	conv	56	56	64	3	3	128	28	28	128	784	128	576	115605504	2502656	46.19313	FALSE
Layer 7	conv	28	28	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 8	conv	28	28	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 9	conv	28	28	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 10	conv	28	28	128	3	3	256	14	14	256	196	256	1152	115605504	2283520	50.62601	FALSE
Layer 11	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 12	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 13	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 14	conv	14	14	256	3	3	512	7	7	512	49	512	2304	115605504	5270528	21.93433	TRUE
Layer 15	conv	7	7	512	3	3	512	7	7	512	49	512	4608	231211008	10440704	22.14515	TRUE
Layer 16	conv	7	7	512	3	3	512	7	7	512	49	512	4608	231211008	10440704	22.14515	TRUE
Layer 17	conv	7	7	512	3	3	512	7	7	512	49	512	4608	231211008	10440704	22.14515	TRUE
Layer 18	fc	7	7	512	7	7	1000	1	1	1000	1	1000	25088	50176000	1E+08	0.499481	TRUE

VGG-16	ImageNet				Single batch size											
		H	W	C	R	S	K	P	Q	K	m (P*Q)	n (K)	k (R*S*C)			
Layer 1	conv	224	224	3	3	3	64	224	224	64	50176	64	27	173408256	18270976	9.490914	TRUE
Layer 2	conv	224	224	64	3	3	64	224	224	64	50176	64	576	3699376128	1.29E+08	28.76698	FALSE
Layer 3	conv	112	112	64	3	3	128	112	112	128	12544	128	576	1849688064	35618816	51.93008	FALSE
Layer 4	conv	112	112	128	3	3	128	112	112	128	12544	128	1152	3699376128	64815104	57.07583	FALSE
Layer 5	conv	56	56	128	3	3	128	56	56	256	3136	256	1152	1849688064	18841600	98.17043	FALSE
Layer 6	conv	56	56	256	3	3	256	56	56	256	3136	256	2304	3699376128	34471936	107.3156	FALSE
Layer 7	conv	56	56	256	3	3	256	56	56	256	3136	256	2304	3699376128	34471936	107.3156	FALSE
Layer 8	conv	28	28	256	3	3	512	28	28	512	784	512	2304	1849688064	13549568	136.5127	FALSE
Layer 9	conv	28	28	512	3	3	512	28	28	512	784	512	4608	3699376128	25493504	145.1105	FALSE
Layer 10	conv	28	28	512	3	3	512	28	28	512	784	512	4608	3699376128	25493504	145.1105	FALSE
Layer 11	conv	14	14	512	3	3	512	14	14	512	196	512	4608	924844032	13451264	68.75518	FALSE
Layer 12	conv	14	14	512	3	3	512	14	14	512	196	512	4608	924844032	13451264	68.75518	FALSE
Layer 13	conv	14	14	512	3	3	512	14	14	512	196	512	4608	924844032	13451264	68.75518	FALSE
Layer 14	fc	7	7	512	7	7	4096	1	1	4096	1	4096	25088	205520896	4.11E+08	0.499858	TRUE
Layer 15	fc	1	1	4096	1	1	4096	1	1	4096	1	4096	4096	33554432	67141632	0.499756	TRUE
Layer 16	fc	1	1	4096	1	1	1000	1	1	1000	1	1000	4096	8192000	16404384	0.499379	TRUE

ResNet-50	ImageNet				Single batch size											
		H	W	C	R	S	K	P	Q	K	m (P*Q)	n (K)	k (R*S*C)	#operations	#bytes	22.9547	
Layer 1	conv	227	227	3	7	7	64	112	112	64	12544	64	147	236027904	10624768	22.21488	TRUE
Layer 2	conv	56	56	64	1	1	64	56	56	64	3136	64	64	25690112	1622016	15.83838	TRUE
Layer 3	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 4	conv	56	56	64	1	1	256	56	56	256	3136	256	64	102760448	4079616	25.18876	FALSE
Layer 5	conv	56	56	256	1	1	64	56	56	64	3136	64	256	102760448	4079616	25.18876	FALSE
Layer 6	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 7	conv	56	56	64	1	1	256	56	56	256	3136	256	64	102760448	4079616	25.18876	FALSE
Layer 8	conv	56	56	256	1	1	64	56	56	64	3136	64	256	102760448	4079616	25.18876	FALSE
Layer 9	conv	56	56	64	3	3	64	56	56	64	3136	64	576	231211008	8175616	28.28056	FALSE
Layer 10	conv	56	56	64	1	1	256	56	56	256	3136	256	64	102760448	4079616	25.18876	FALSE
Layer 11	conv	56	56	256	1	1	128	56	56	128	3136	128	256	205520896	4947968	41.53642	FALSE
Layer 12	conv	56	56	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 13	conv	28	28	128	1	1	512	28	28	512	784	512	128	102760448	2269184	45.2852	FALSE
Layer 14	conv	28	28	512	1	1	128	28	28	128	784	128	512	102760448	2269184	45.2852	FALSE
Layer 15	conv	28	28	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 16	conv	28	28	128	1	1	512	28	28	512	784	512	128	102760448	2269184	45.2852	FALSE
Layer 17	conv	28	28	512	1	1	128	28	28	128	784	128	512	102760448	2269184	45.2852	FALSE
Layer 18	conv	28	28	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 19	conv	28	28	128	1	1	512	28	28	512	784	512	128	102760448	2269184	45.2852	FALSE
Layer 20	conv	28	28	512	1	1	128	28	28	128	784	128	512	102760448	2269184	45.2852	FALSE
Layer 21	conv	28	28	128	3	3	128	28	28	128	784	128	1152	231211008	4603904	50.22064	FALSE
Layer 22	conv	28	28	128	1	1	512	28	28	512	784	512	128	102760448	2269184	45.2852	FALSE
Layer 23	conv	28	28	512	1	1	256	28	28	256	784	256	512	205520896	2932736	70.07821	FALSE
Layer 24	conv	28	28	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 25	conv	14	14	256	1	1	1024	14	14	1024	196	1024	256	102760448	2052096	50.07585	FALSE
Layer 26	conv	14	14	1024	1	1	256	14	14	256	196	256	1024	102760448	2052096	50.07585	FALSE
Layer 27	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 28	conv	14	14	256	1	1	1024	14	14	1024	196	1024	256	102760448	2052096	50.07585	FALSE

Layer 29	conv	14	14	1024	1	1	256	14	14	256	196	256	1024	102760448	2052096	50.07585	FALSE
Layer 30	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 31	conv	14	14	256	1	1	1024	14	14	1024	196	1024	256	102760448	2052096	50.07585	FALSE
Layer 32	conv	14	14	1024	1	1	256	14	14	256	196	256	1024	102760448	2052096	50.07585	FALSE
Layer 33	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 34	conv	14	14	256	1	1	1024	14	14	1024	196	1024	256	102760448	2052096	50.07585	FALSE
Layer 35	conv	14	14	1024	1	1	256	14	14	256	196	256	1024	102760448	2052096	50.07585	FALSE
Layer 36	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 37	conv	14	14	256	1	1	1024	14	14	1024	196	1024	256	102760448	2052096	50.07585	FALSE
Layer 38	conv	14	14	1024	1	1	256	14	14	256	196	256	1024	102760448	2052096	50.07585	FALSE
Layer 39	conv	14	14	256	3	3	256	14	14	256	196	256	2304	231211008	4366336	52.9531	FALSE
Layer 40	conv	14	14	256	1	1	1024	14	14	1024	196	1024	256	102760448	2052096	50.07585	FALSE
Layer 41	conv	14	14	1024	1	1	512	14	14	512	196	512	1024	205520896	3301376	62.2531	FALSE
Layer 42	conv	14	14	512	3	3	512	7	7	512	49	512	4608	231211008	10440704	22.14515	TRUE
Layer 43	conv	7	7	512	1	1	2048	7	7	2048	49	2048	512	102760448	4696064	21.88225	TRUE
Layer 44	conv	7	7	2048	1	1	512	7	7	512	49	512	2048	102760448	4696064	21.88225	TRUE
Layer 45	conv	7	7	512	3	3	512	7	7	512	49	512	4608	231211008	10440704	22.14515	TRUE
Layer 46	conv	7	7	512	1	1	2048	7	7	2048	49	2048	512	102760448	4696064	21.88225	TRUE
Layer 47	conv	7	7	2048	1	1	512	7	7	512	49	512	2048	102760448	4696064	21.88225	TRUE
Layer 48	conv	7	7	512	3	3	512	7	7	512	49	512	4608	231211008	10440704	22.14515	TRUE
Layer 49	conv	7	7	512	1	1	2048	7	7	2048	49	2048	512	102760448	4696064	21.88225	TRUE
Layer 50	fc	1	1	2048	1	1	1000	1	1	1000	1	1000	2048	4096000	8204192	0.499257	TRUE