

layer name	layer depth	output size	Small	Medium	Large	Extra large
Downsampling	1	$W/2 \times H/2$	$1 \times 1, 64$ $3 \times 3, 64$ stride 2 $\times 1$	$1 \times 1, 64$ $3 \times 3, 64$ stride 2 $\times 1$	$1 \times 1, 64$ $3 \times 3, 64$ stride 2 $\times 1$	$1 \times 1, 64$ $3 \times 3, 64$ stride 2 $\times 1$
Downsampling	2	$W/4 \times H/4$	$1 \times 1, 96$ $3 \times 3, 96$ stride 2 $\times 1$	$1 \times 1, 128$ $3 \times 3, 128$ stride 2 $\times 1$	$1 \times 1, 160$ $3 \times 3, 160$ stride 2 $\times 1$	$1 \times 1, 200$ $3 \times 3, 200$ stride 2 $\times 1$
Downsampling	3	$W/8 \times H/8$	$1 \times 1, 192$ $3 \times 3, 192$ stride 2 $\times 1$	$1 \times 1, 256$ $3 \times 3, 256$ stride 2 $\times 1$	$1 \times 1, 320$ $3 \times 3, 320$ stride 2 $\times 1$	$1 \times 1, 400$ $3 \times 3, 400$ stride 2 $\times 1$
Downsampling	4	$W/16 \times H/16$	$1 \times 1, 384$ $3 \times 3, 384$ stride 2 $\times 1$	$1 \times 1, 512$ $3 \times 3, 512$ stride 2 $\times 1$	$1 \times 1, 640$ $3 \times 3, 640$ stride 2 $\times 1$	$1 \times 1, 800$ $3 \times 3, 800$ stride 2 $\times 1$
Stream1	3-6	$W/4 \times H/4$	$1 \times 1, 96$ $3 \times 3, 96$ SSE $\times 4$	$1 \times 1, 128$ $3 \times 3, 128$ SSE $\times 4$	$1 \times 1, 160$ $3 \times 3, 160$ SSE $\times 4$	$1 \times 1, 200$ $3 \times 3, 200$ SSE $\times 4$
Stream1-Downsampling	8	$W/8 \times H/8$	$1 \times 1, 192$ $3 \times 3, 192$ stride 2 $\times 1$	$1 \times 1, 256$ $3 \times 3, 256$ stride 2 $\times 1$	$1 \times 1, 320$ $3 \times 3, 320$ stride 2 $\times 1$	$1 \times 1, 400$ $3 \times 3, 400$ stride 2 $\times 1$
Stream2	4-8	$W/8 \times H/8$	$1 \times 1, 192$ $3 \times 3, 192$ SSE $\times 5$	$1 \times 1, 256$ $3 \times 3, 256$ SSE $\times 5$	$1 \times 1, 320$ $3 \times 3, 320$ SSE $\times 5$	$1 \times 1, 400$ $3 \times 3, 400$ SSE $\times 5$
Stream2-Fusion	9	$W/16 \times H/16$	$1 \times 1, 384$ $3 \times 3, 384$ stride 2 $\times 1$	$1 \times 1, 512$ $3 \times 3, 512$ stride 2 $\times 1$	$1 \times 1, 640$ $3 \times 3, 640$ stride 2 $\times 1$	$1 \times 1, 800$ $3 \times 3, 800$ stride 2 $\times 1$
Stream3	5-9	$W/16 \times H/16$	$1 \times 1, 384$ $3 \times 3, 384$ SSE $\times 5$	$1 \times 1, 512$ $3 \times 3, 512$ SSE $\times 5$	$1 \times 1, 640$ $3 \times 3, 640$ SSE $\times 5$	$1 \times 1, 800$ $3 \times 3, 800$ SSE $\times 5$
Stream3-Fusion	10	$W/16 \times H/16$	$1 \times 1, 384$ $3 \times 3, 384$ SSE $\times 1$	$1 \times 1, 512$ $3 \times 3, 512$ SSE $\times 1$	$1 \times 1, 640$ $3 \times 3, 640$ SSE $\times 1$	$1 \times 1, 800$ $3 \times 3, 800$ SSE $\times 1$
Downsampling	11	$W/32 \times H/32$	$1 \times 1, 1280$ $3 \times 3, 1280$ stride 2 $\times 1$	$1 \times 1, 2048$ $3 \times 3, 2048$ stride 2 $\times 1$	$1 \times 1, 2560$ $3 \times 3, 2560$ stride 2 $\times 1$	$1 \times 1, 3200$ $3 \times 3, 3200$ stride 2 $\times 1$
Final	12	1×1	average pool, 1000-d fc, softmax			

Table A1: Specification of ParNet models used for ImageNet classification: ParNet-S, ParNet-M, ParNet-L, and ParNet-XL.

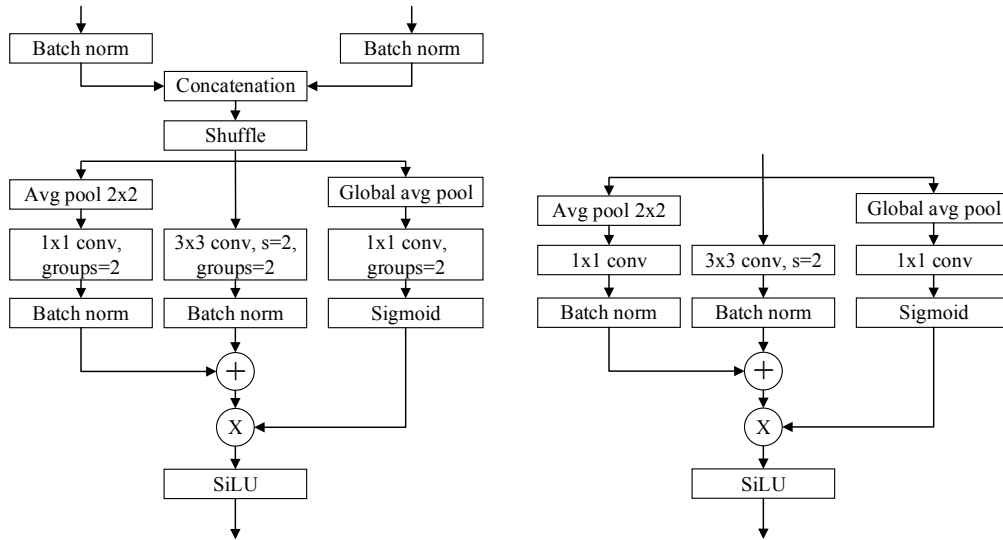


Figure A1: Schematic representation of the Fusion (left) and Downsampling (right) blocks used in ParNet.