

layer	video encoder		audio encoder		
	# filters	kernel	# filters	kernel	stride
1	128	5×5	64	5×5	2×2
2	128	5×5	64	4×4	1×1
3	256	3×3	128	4×4	2×2
4	256	3×3	128	2×2	2×1
5	512	3×3	128	2×2	2×1
6	512	3×3			