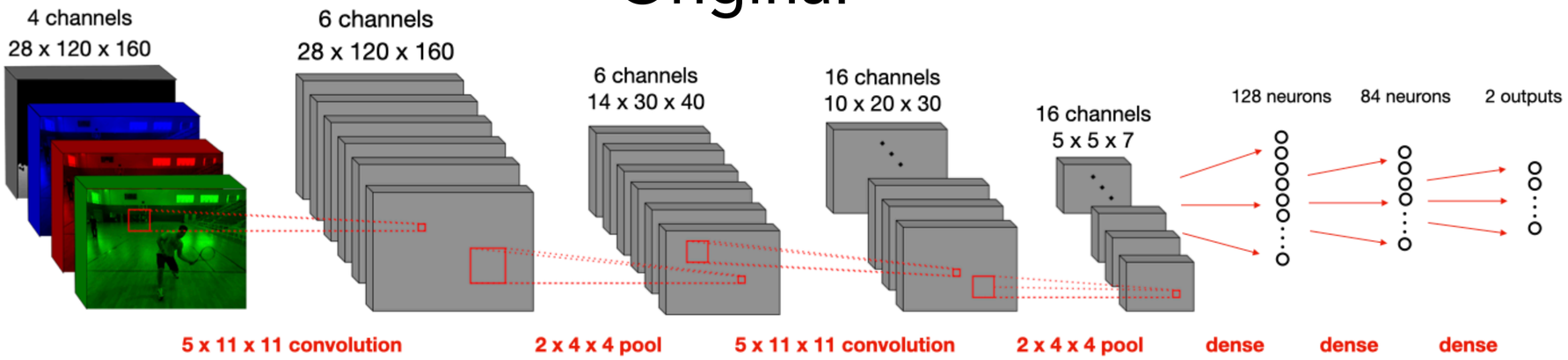
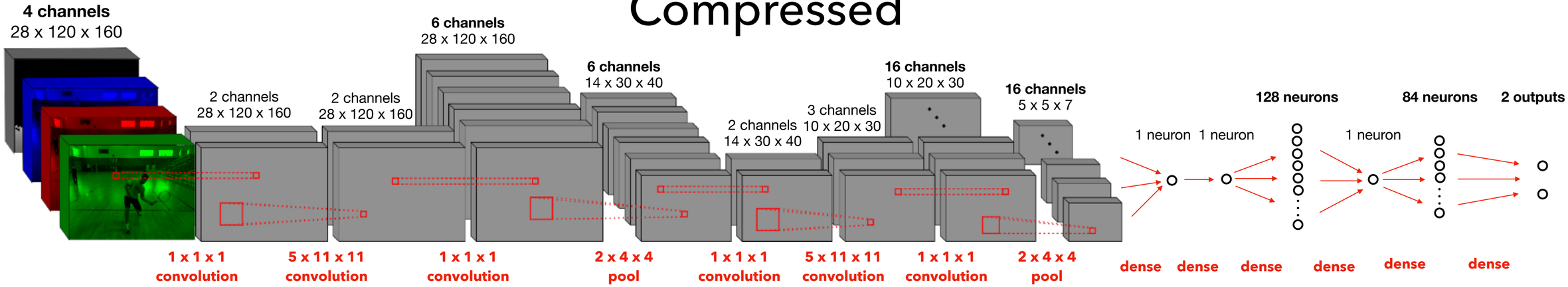
Method 2 results

THETIS

Original



Compressed



The channels marked with **bold** correspond to the layer in the original model

Method 2 results

Overall speed-ups

	Theoretical	Observed	Storage improvements	Accuracy

MNIST — × 3.45 | × 0.58 | × 4.62

Original	Compressed		Drop
	Before fine-tuning	After fine-tuning	
98.66 %	66.12 %	98.19 %	-0.47 %

THETIS — × 6.12 | × 1.34 | × 33.13

Original	Compressed		Drop
	Before fine-tuning	After fine-tuning	
92 %	68 %	90 %	-2

VGG-16 — × 5.55 | × 1.82 | × 6.07 | —

The VGG-16 model is well-known and have been downloaded pre-trained