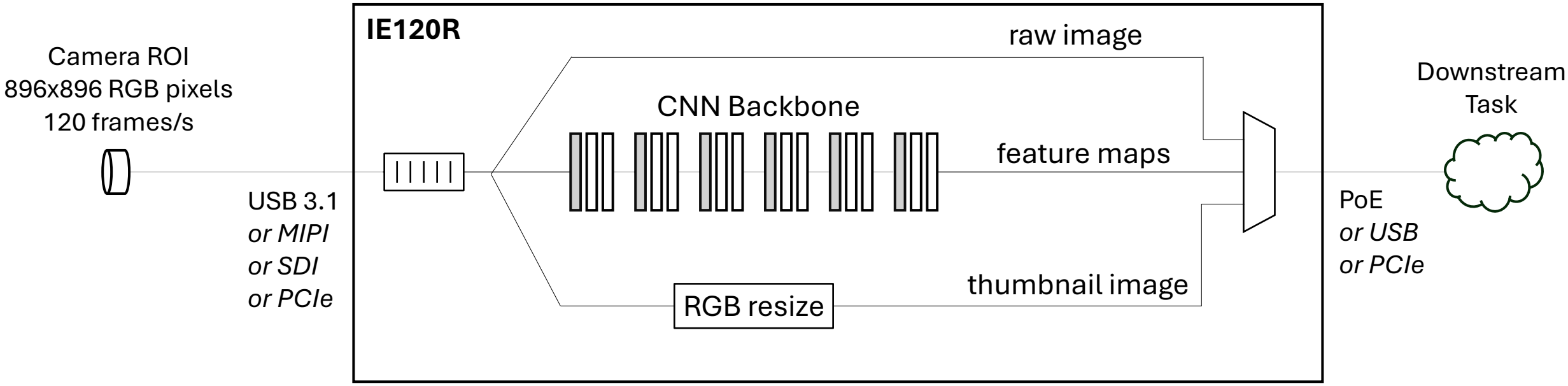
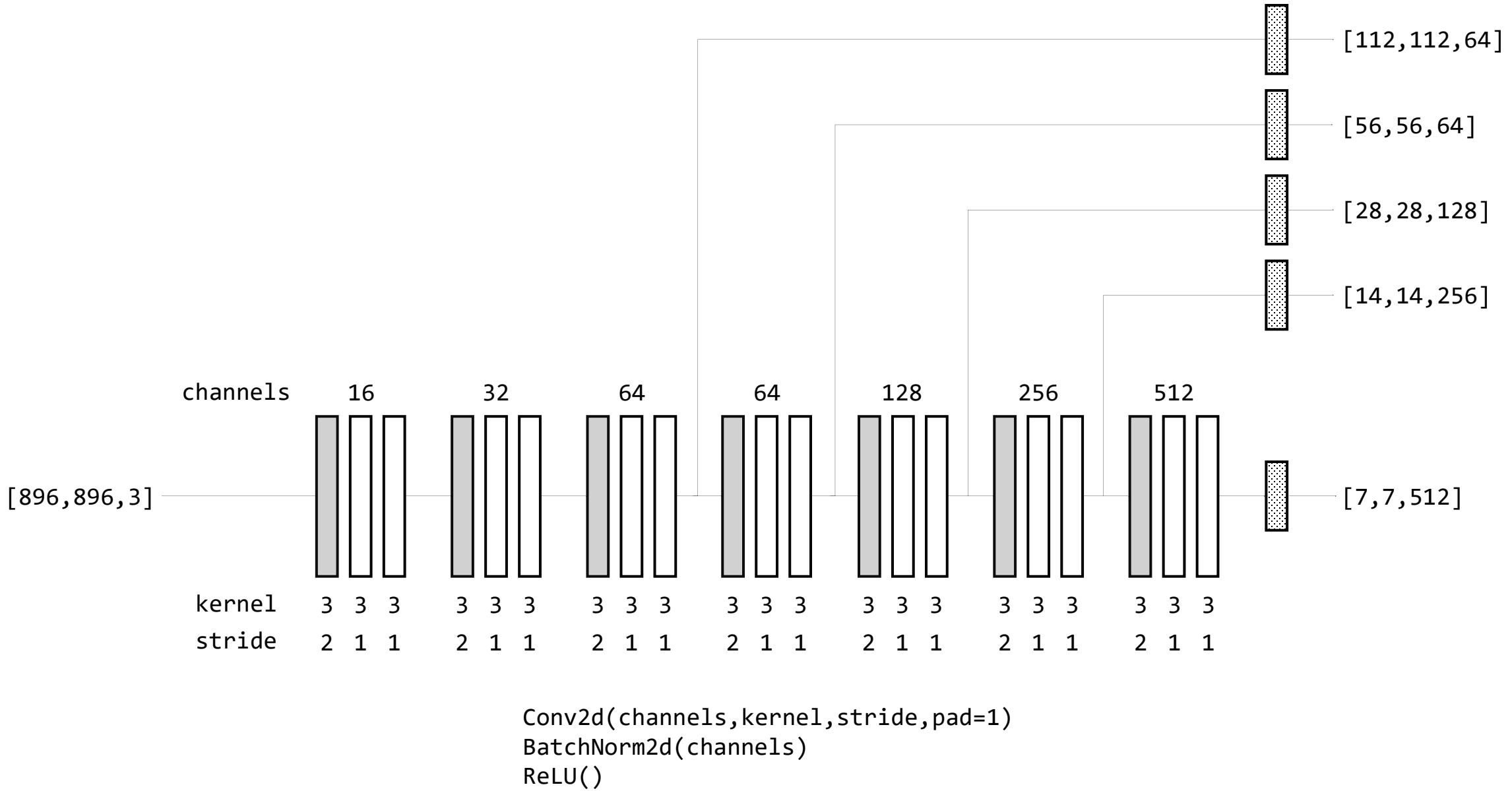
IE120R Single Chip Image Encoder



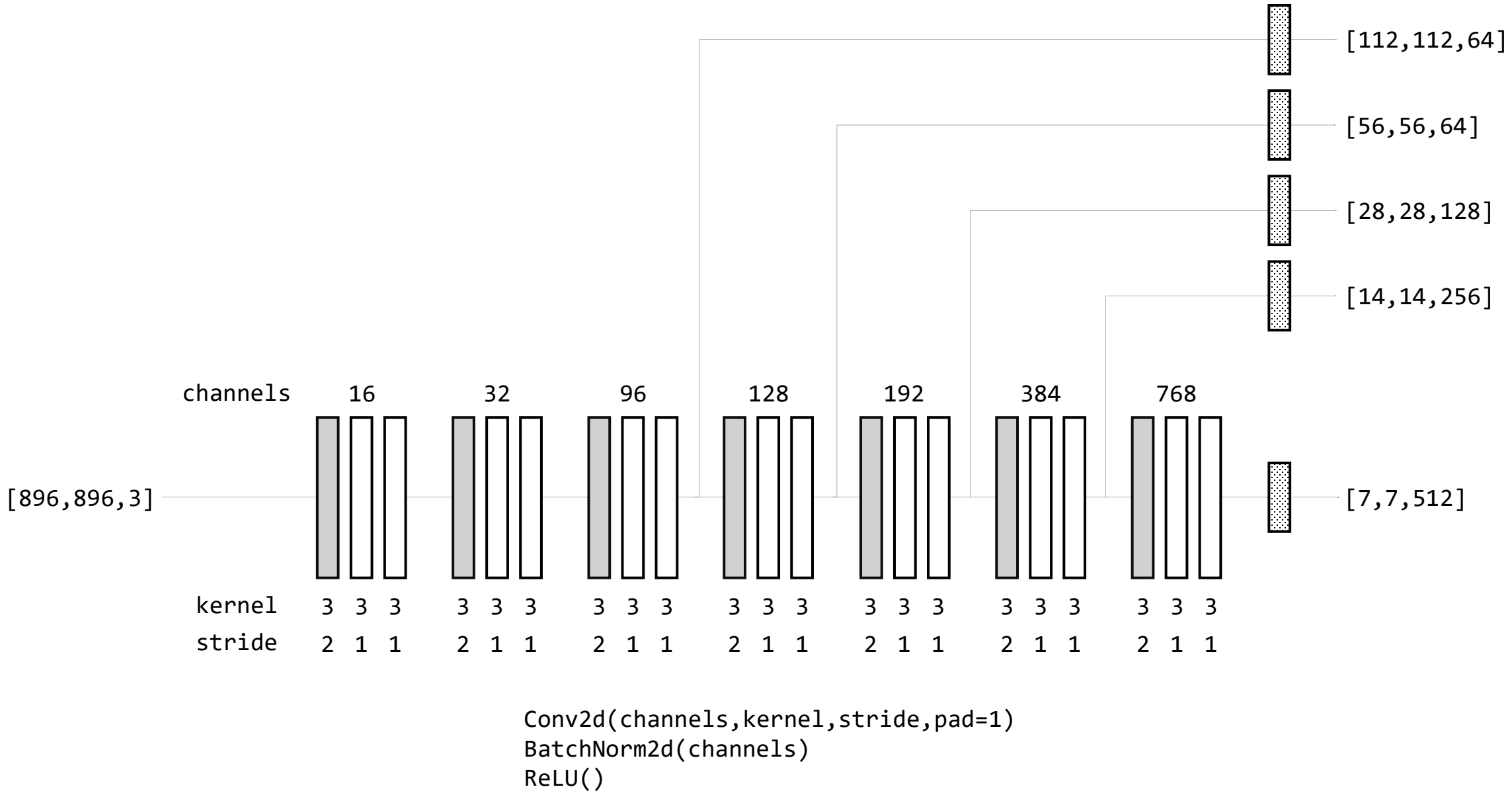
IE120R-medium Data Sheet

Compute Fabric	Agilex AGF 027
PyTorch Model	VGG-like CNN backbone, 22 layers, 18.65M weights (FP8)
Input Shape	RGB image (FP32) → 896x896 pixel ROI
Output Shapes	Feature maps (FP32) → [7,7,512] + [14,14,256] + [28,28,128] + [56,56,64] + [112,112,64]
Throughput	120 frames/s
Latency	<2 ms
Power	<15 W
Camera Interface	USB 3.1 5Gb/s, USB3 Vision Standard
Downstream Interface	PoE 12W, ARP, UDP, RTP/RTSP, HTTPS

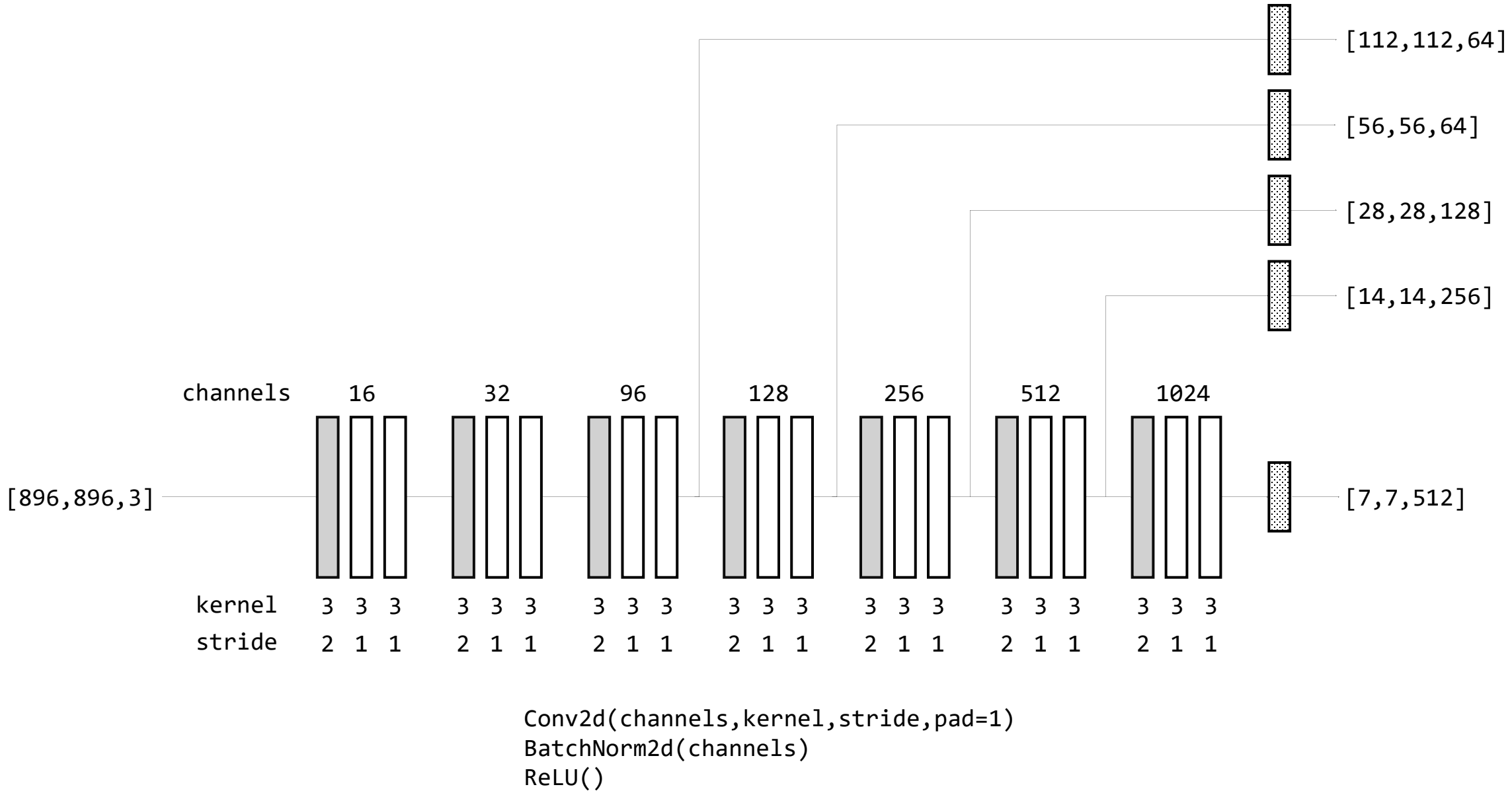
IE120R-small



IE120R-medium



IE120R-large



Resnet-18

Skip connections require additional intermediate state storage and global interconnect

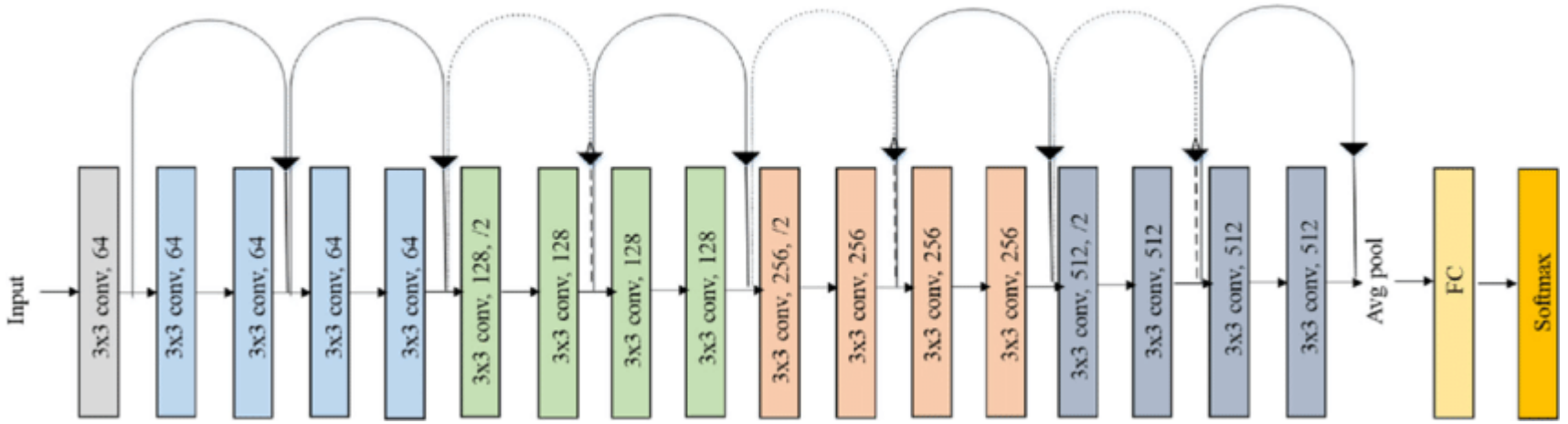
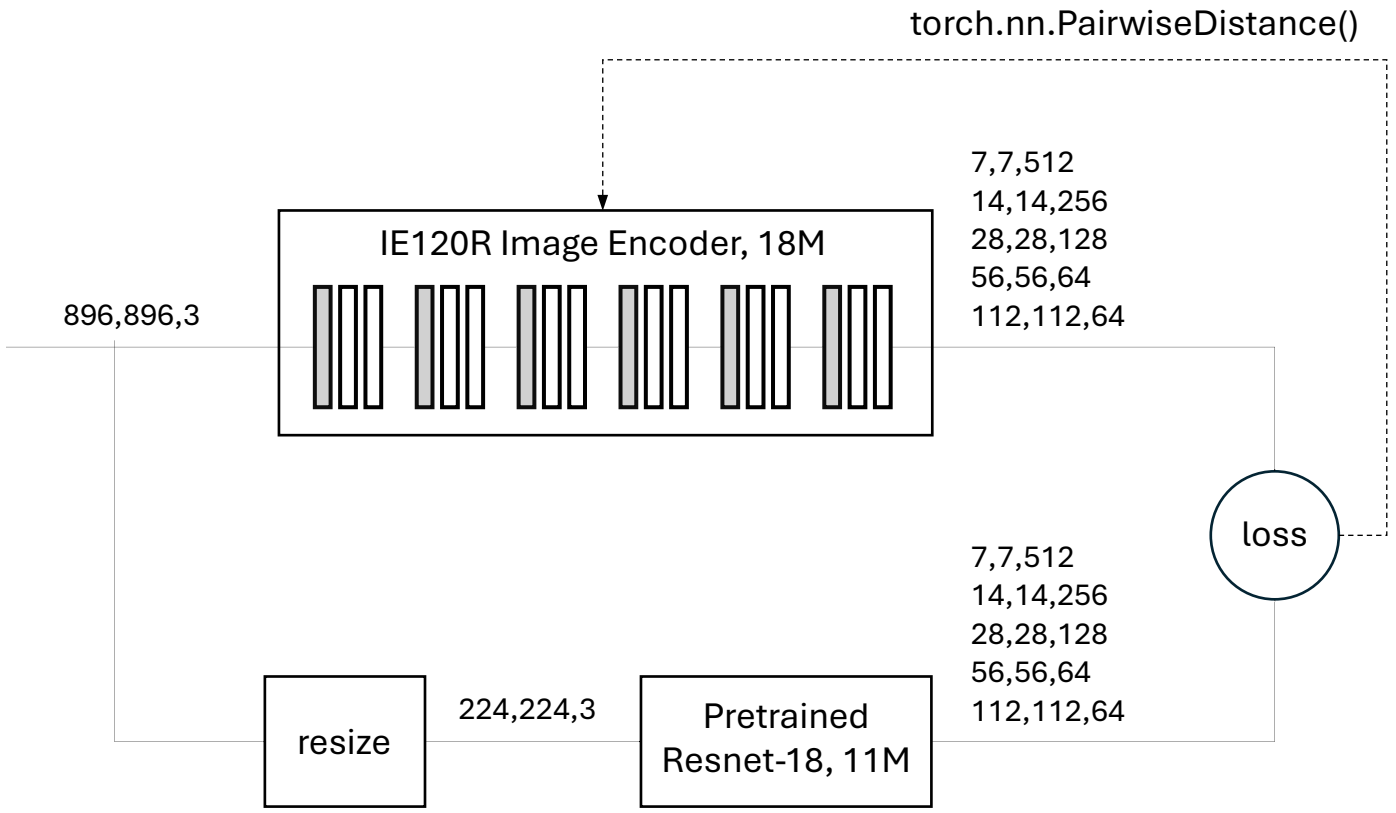


Image Encoder Pretraining

unlabeled images



https://github.com/siliconperception/models/blob/main/pretrain/resnet.py

Image Encoder Testing

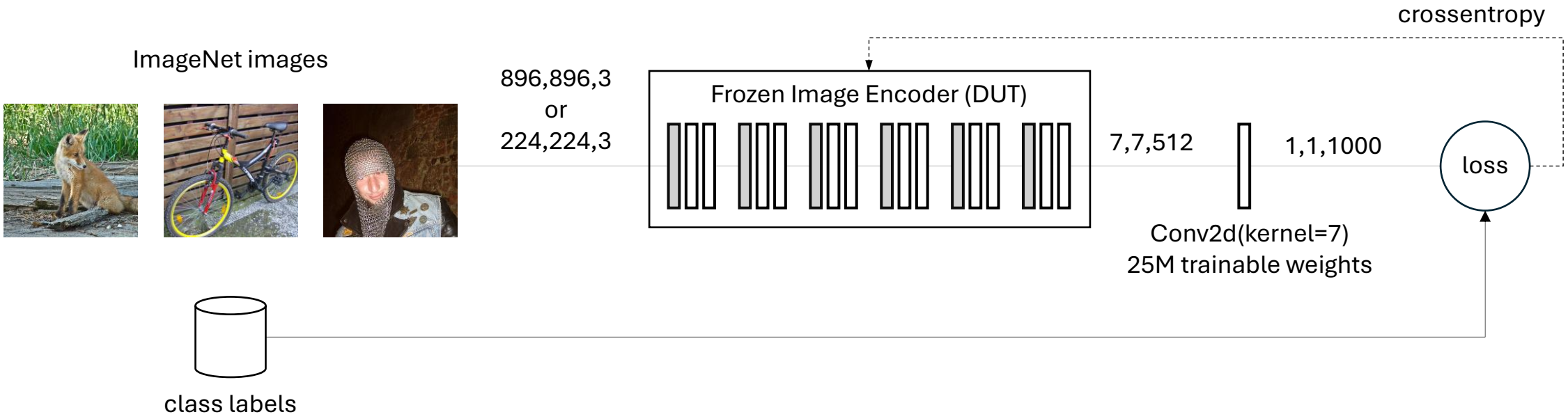


Image Encoder Pretraining Results

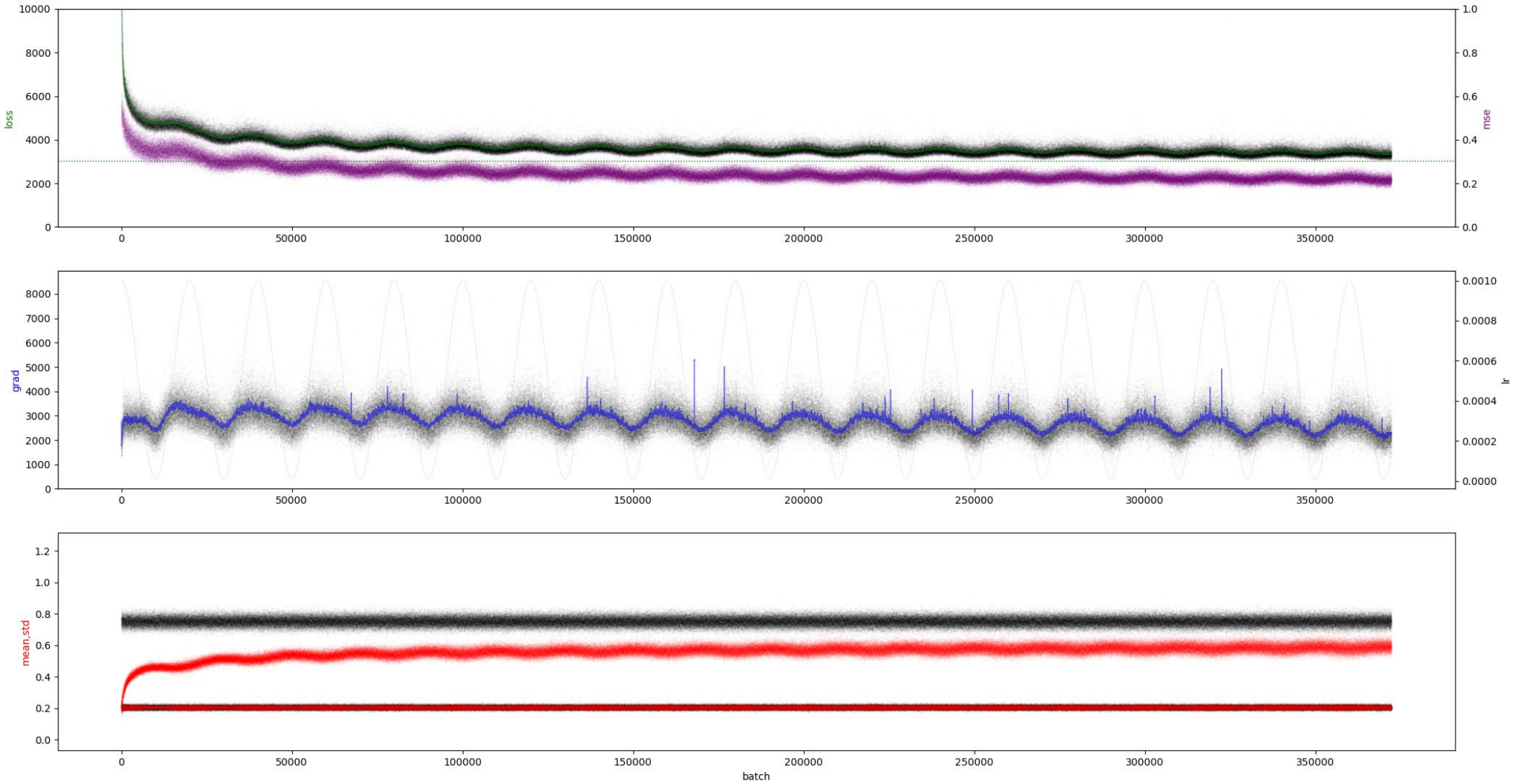


Image Encoder Test Results

Model (DUT)	Weights	MAC	Input	Accuracy
Resnet-18	11.18M	1.81G	[224,224,3]	47.3%
IE120R-small	8.34M	4.66G	[896,896,3]	41.2%
IE120R-medium	18.65M	8.04G	[896,896,3]	47.4%
IE120R-large	32.17M	8.12G	[896,896,3]	49.8%

1. Knowledge distillation using pairwise distance is effective
2. Distilled model accuracy is proportional to weight capacity