

# 学习计划

汇报人：焦强

# 一，关于本周学习

这星期看了李宏毅视频，跟着做了几个例题作业。  
做UG902的实验。  
然后就是看论文。



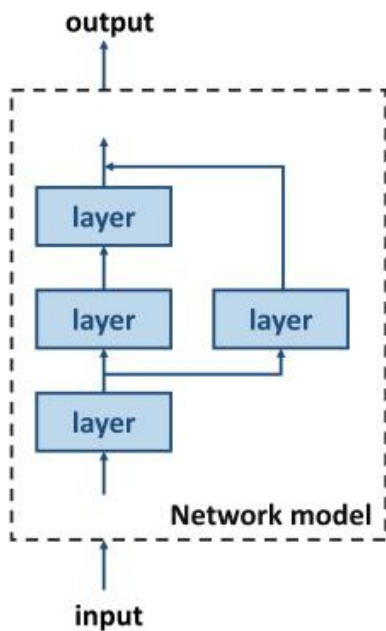
两篇综述论文，我总结了下。  
神经网络加速器的设计目标包括以下两个方面：  
高速(高吞吐量和低延迟)和高能效。

高吞吐量：

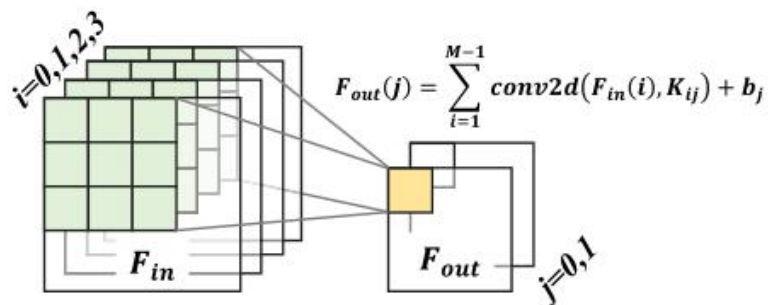
- 1，减小计算量或者每个计算单元的大小
- 2，增加工作频率来提高峰值性能
- 3，合理的并行实现以及数据交换

高能效：

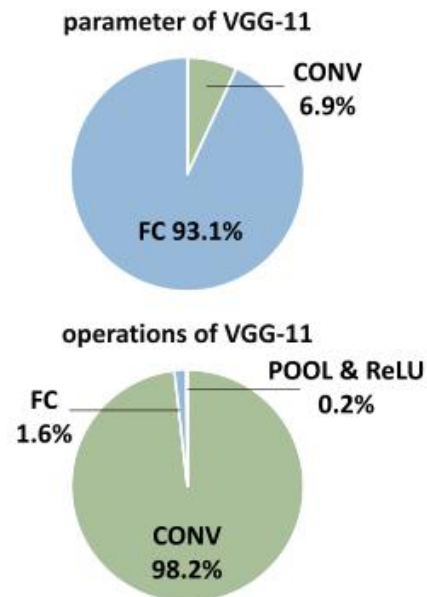
- 1：计算
- 2：内存访问



(a)



(b)



(c)

Fig. 1. (a) Computation graph of a neural network model. (b) CONV and FC layers in an NN model. (c) CONV and FC layers dominate the computation and parameter of a typical NN model: VGG11.

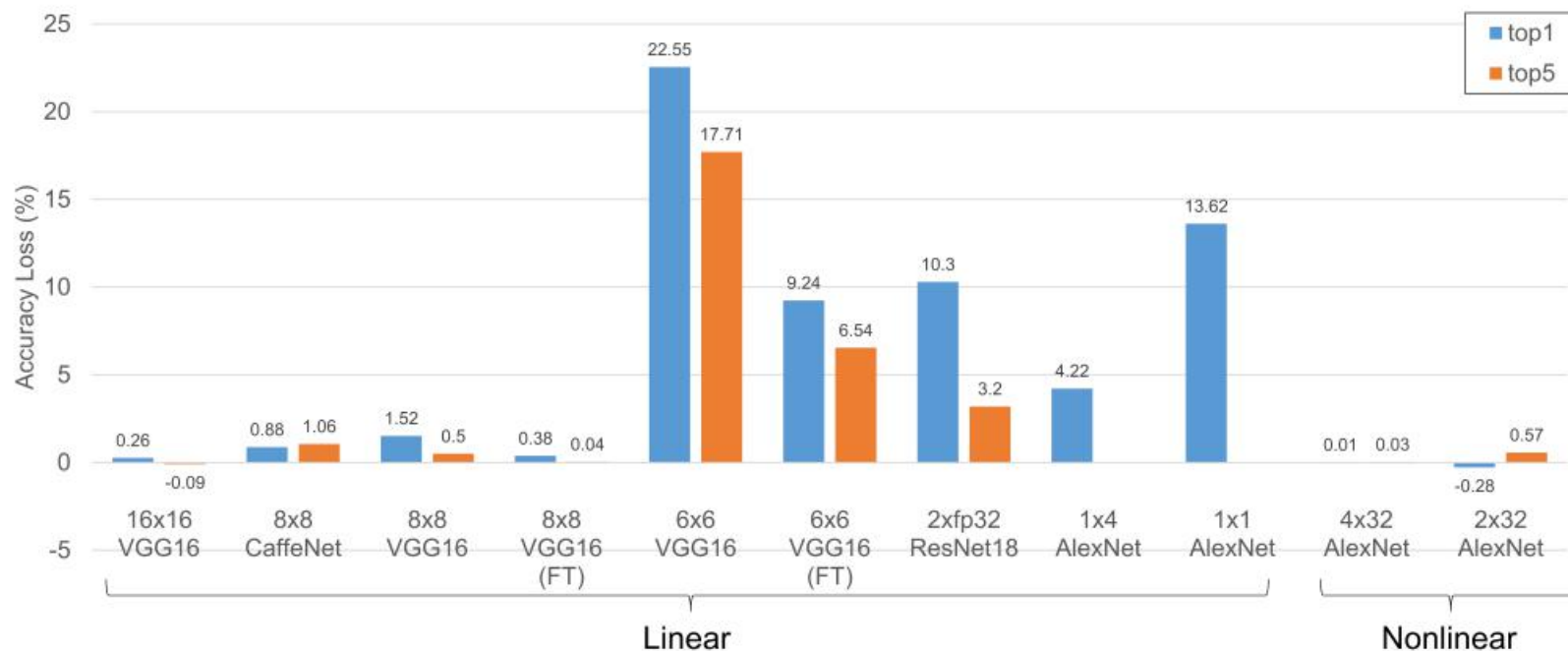


Fig. 3. Comparison between different quantization methods from References [17, 20, 29, 49, 82, 83]. The quantization configuration is expressed as (weight bit-width) × (activation bit-width). The "(FT)" denotes that the network is fine-tuned after a linear quantization.

Table 3. FPGA Resource Consumption Comparison for Multiplier and Adder with Different Types of Data

	Xilinx Logic				Xilinx DSP			Altera DSP	
	Multiplier		Adder		Multiply and add			Multiply and add	
	LUT	FF	LUT	FF	LUT	FF	DSP	ALM	DSP
fp32	708	858	430	749	800	1284	2	1	1
fp16	221	303	211	337	451	686	1	213	1
fixed32	1112	1143	32	32	111	64	4	64	3
fixed16	289	301	16	16	0	0	1	0	1
fixed8	75	80	8	8	0	0	1	0	1
fixed4	17	20	4	4	0	0	1	0	1

# 年前的学习计划

- 1,李宏毅视频看完。
- 2, 常见的CNN网络跑一遍。
- 3, 把xilinx官方的手册看几遍, 熟悉Vivado HLS工具的使用。

# 下周学习计划

- 1,李宏毅视频看完CNN。
- 2, 把xilinx的ug902手册的实验往后做
- 3, 再找论文看。