Reviewer: 1

Comments to the Author

The writing is still very poor. This paper will need a lot of copy editing. We can start with the title:

A Survey of FPGA Based Neural Network Inference Accelerator

should be

A Survey of FPGA-Based Neural Network Inference Accelerators

根据意见修改了题目，顺便把所有的FPGA based改成了FPGA-based

Following up with another response to Reviewer 3. If "Most of the techniques introduced are suitable to both CNN and RNN or other state-of-the-art networks" then it would be good to discuss this some more as the paper seems to be more about CNNs.

根据意见，在2.1的第四段做了说明，RNN的组件相比CNN少了CONV层，虽然用CNN举例，但是这些分析对RNN同样适用。同时在5.2的开始强调了类似的方法对RNN也适用，但是因为CONV层的情况更复杂，所以用CNN为例子说明问题

Section 3:

This section was changed to address comments of Reviewer 3, but I now find it worse that it was before. Before, there was at least some textual discussion that made more sense. Now there are just equations with little to no discusson about the equations and what they are showing. You have to use Table 1 to decipher what the equations mean. Table 1 is still a good idea, but you should keep the textual descriptions and formalize them with the equations.

根据意见修改了第三段中的一些说明。我觉得reviewer之前没理解，有一部分原因是我说明的变量没在公式里出现。比如频率和并行度没有在公式1里面出现。所以现在加上了。另外，说明的时候一个变量如果有符号的话，也同时加上了名称和变量名，方便理解。另一方面主要是对公式4解释的时候，在变量表里面换了变量名，但是文字里忘了改了，所以造成比较难懂。

由于这一段有承上启下的作用，所以也在这一段说明可能采用的优化方法的时候，介绍了相关的技术会在后面的哪个章节介绍。

p 11.16: "... there is still 1.3 1.5x gap ..." Should this be "... there is still 1.3 - 1.5x gap ..."?

根据意见改正

p 11.17: Fig. 7. You have not addressed my concern. You make comments on BRAM/DSP and Logic/DSP. You still do not comment on the Logic/BRAM chart. I previously said that if you do not comment about it in the text, it should be removed. However, I would say that you need it and the text should make an observation about what that chart shows.

原本没有解释三个图中的最后一个，现在加了一个解释，就是设计对LUT和BRAM的需求没有很确定的倾向性，因为有的设计会用LUT来搭建更多的计算单元寻求高并行度，而有的设计则倾向于只用DSP来寻求高的计算频率。

Reviewer: 2

Comments to the Author

(There are no comments.)

Reviewer: 3

Comments to the Author

Do include a citation to Stylianos I. Venieris, Alexandros Kouris, Christos-Savvas Bouganis: Toolflows for Mapping Convolutional Neural Networks on FPGAs: A Survey and Future Directions. ACM Comput. Surv. 51(3): 56:1-56:39 (2018)

已添加引用，在8.1的第一段。主要说明引用的这篇文章在toolflow方面的介绍更详细。而本文的这一部分从另一个角度进行了一点补充。

另外：引用了deming老师的几篇文章，分布在全文各处；根据新的引用修改了图6中的引用，也整理了一下表3中的不同设计的顺序，按照位宽和性能排了序