Dear Reviewers:

Thank you for the reviewers’ comments. We have studied comments carefully and have corrected. The responses are as follows:

**Reviewer #1:**

1. This paper need a major revision. The current expression of the novel ideas are not clear**.** Quantization is a commonly used technique for FPGA based implementations and multi-PE is also not totally novel. I believe the systematic design of the paper should be a good one, but it is not convincible enough now.

**Response:**

**B**inarized **n**eural **n**etworks (BNN) is a generic term, which contains many kinds of neural networks, such as binarized recurrent neural networks and binarized convolutional neural networks. The BinaryNet model used in the paper only contains convolutional layers and FC layers. It will be more accurate to call it a BCNN model. For clarity, we have modified Section 1 with more accurate abbreviations. At the beginning of Section 2, to maintain the custom of expression, we have stated that BNN is equivalent to BCNN in subsequent parts of the paper.