Response to Reviewer 3 Comments

Dear reviewer,

Thank you very much for your recognition and your valuable comments. This time, I modify my manuscript according to your suggestion. At the same time, my classmates helped me correct some of the English writing errors and description errors in the manuscript.

For your suggestions, I explained each point in detail as follows, and revised the corresponding content in the manuscript. At the same time, we enriched our references and results. I hope that you are satisfied with my processing results.

**Point 1:** Why adopt this method over classical compression techniques? How it compares to plain Huffman (or similar, like correction codes)?

**Response 1:** We add a summary of coding techniques in the introduction chapter of the new manuscript. (on lines 61-71 in the new manuscript) The biggest difference between this method and the traditional method is the idea of using neural network. At the same time, we consider the compression problem from the mathematical distribution characteristics of the data. We use neural network to learn the distribution characteristics of the data, and reconstruct the data by using the learned distribution. Compared to the encoding method, our method can achieve a higher compression ratio. At the same time our experiments prove that the error of our method is low enough.

**Point 2:** What is exactly the power consumption of training/run the network on the node? How it compares to other similar approaches?

**Response 2:** We did not deploy our method to the sensor node for power testing. We calculated the computational complexity of our method and tested the computational performance of different microcontrollers. The actual power consumption is affected by many factors, we can not give a certain value, we can only give a data description from a theoretical point of view. (on lines 610-620 in the new manuscript)

Table 4 shows the comparison of our method with other similar methods. We compare the computational complexity, algorithm parameters, compression time and other aspects to prove the advantages of our algorithm. (on lines 604-608 in the new manuscript)

**Point 3:** In which degree the technique impacts the data overhead due to weigths (or topology, or any parameters) tranfers between the server and the nodes?

**Response 3:** We give the application of our method in the manuscript, we propose two solutions. (on lines 621-633 in the new manuscript) Servers and nodes only need to transfer weights once. At the same time, the weight of our method is also very small.

**Point 4:** How would you implement this on a real WSN? By the paper, it seems targeted to application level, but data compression and aggregation present some advantages when cross layered (e.g. routing level).

**Response 4:** We give the application of our method in the manuscript, we propose two solutions. (on lines 621-633 in the new manuscript)
As you said, our method is at the application level and we don't consider the routing level. But the transfer learning ability of our method also allows our method to have similar utility for data aggregation. The routing level is also the next issue that we need to consider. Thank you for your suggestion.

**Point 5:** One of the good results is the neuron (model) pruning, however there is no comparison with other ANN (and CNN) compression/pruning methods. How your method compares with (for example) plain SqueezeNet usage?

**Response 5:** We compared the pruning results of our neuron pruning method with other common methods at the same pruning rate, the comparison results are shown in Figure 12. (on lines 660-676 in the new manuscript)
Sorry, we have not used SqueezeNet and have no experience with SqueezeNet for data compression, so we have not compared our method with SqueezeNet.

**Point 6:** In which scenario the method would apply? Usually nodes have a limited well known number of sensors, and a well known range of measurement. In this classical scenario, would your method present any advantage over classical methods?

**Response 6:** Our method are application level and are suitable for most sensing situations. But our method needs as much data as possible to train, so our method hopes to the train sensing data as accurately as possible. The performance of our method is not limited by the number of sensors and the range of measurement. In the same environment, we believe that our approach can achieve higher compression performance than traditional methods.