Response to Reviewer 2 Comments

Point 1: Authors should pay attention to the writing of the paper. There are multiples typos such as: "More powerful works is directly apply…"

Response 1: We were really sorry for our careless mistakes. Thank you for your reminder. The original "More powerful works is directly apply" is modified as "More powerful works are directly applying" **(p. 1, line 23)**. We tried our best to improve the manuscript. And here we did not list all changes but marked in red in the revised paper. We appreciate for your warm work earnestly and hope that the correction will meet with approval.

Point 2: The introduction of this paper needs to be revised extensively so others can really understand authors' motivation and background of this research.

Response 2: Thank you very much for your valuable advice. We have revised the introduction extensively. We pruned some unnecessary contents, modified some sentences that may cause ambiguity, and optimized the structure to more directly explain our motivation and the background of this research: we generalize CNNs directly to the triangle meshes, which provide more efficient and flexible representation of 3D shapes.
All modifications are marked in red in the revised manuscript. **(p. 1-2)**

Point 3: There is no explanation of notations in some equations, such as Eq. (1), what are the meaning of "b,c,d"? Authors should explain.

Response 3: Thank you very much for your correction. In Eq. (1), the symbol and "b,c,d" are really inappropriate. We change "[b, c, d] →= [d, b, c]" to "[b, c, d] → [d, b, c]" and supplement explanation: "where b, c, d are the indices of adjacent faces, and after sorting, the original unordered face indexes become the ordered on the right." **(p. 5, lines 145-146)**

Point 4: There is no explanation in the legend of figures, such as in Figure (6), what are top figures and bottom figures?

Response 4: Thank you for your kind suggestion. For Figure (6), we add some descriptions. Top figures and bottom figures are the segmentation visualization of a vase and a alien of the COSEG dataset. **(p. 9)**

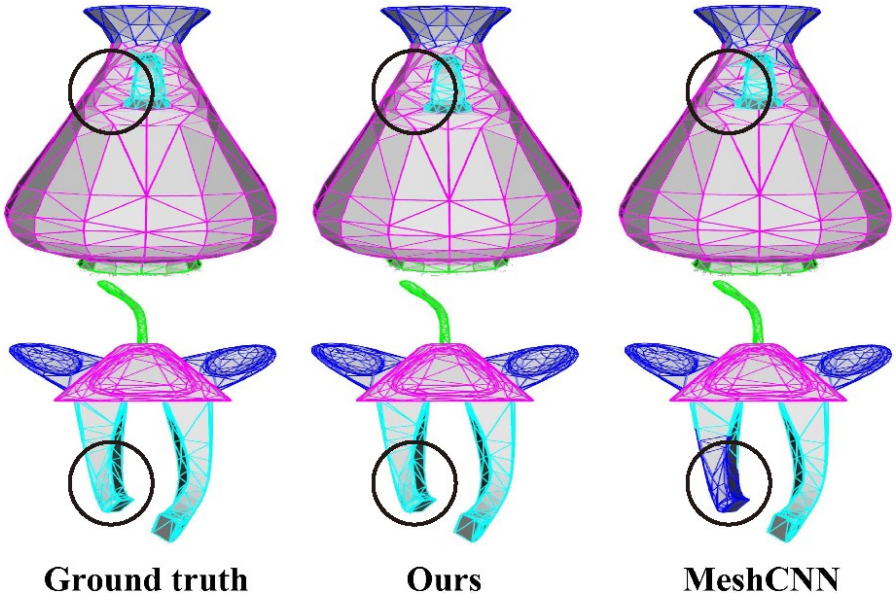

Figure 6. Comparison of qualitative segmentation results with MeshCNN on a vase (top) and a alien (bottom) of the COSEG dataset.

Since a picture has been added, the serial number of this picture should be 7. In order to facilitate you to view, the original picture number (6) is still used.

Point 5: Figure 5 is still not clear to show how the pooling and unpooling works. What are the "l,I,m,n"? In Eq. (8), there are only "l" and "i", so after the unpooling, where are "m,n"?

Response 5: There's really some confusion and it will cause some ambiguity. In fact, Figure 5 show that the pooling is the mesh simplification of face collapse and unpooling is its inverse operation. "l,i,m,n" are the indexes of faces and "m, n" are calculated in the same way as " l", which is the mean of its two adjacent green faces.

In order to eliminate ambiguity and give readers a better understanding, we have added some description to the legend in Figure 5 and supplement above explanation in Eq. (7) and Eq. (8). **(p. 7)**

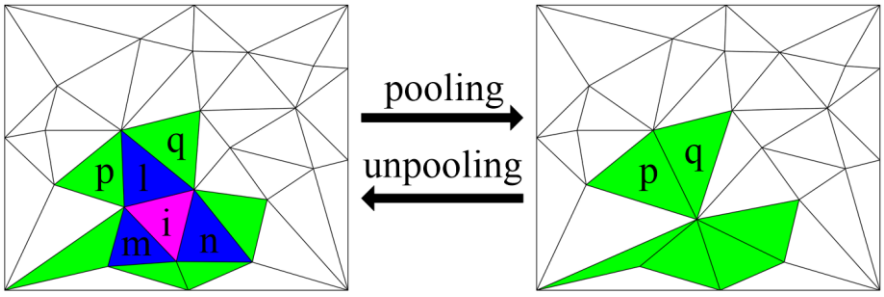

Figure 5. Pooling and unpooling. The purple part is the center face, which together with the blue faces makes up the pooling faces. Similar to the mesh simplification of face collaps, after pooling they will disappear.

Point 6: In Figure 4, what are the blue and orange parts?

Response 6: The orange parts are the center faces, which together with the blue faces constitutes the convolutional receptive fields of the center faces. We have added this description. **(p. 5)**

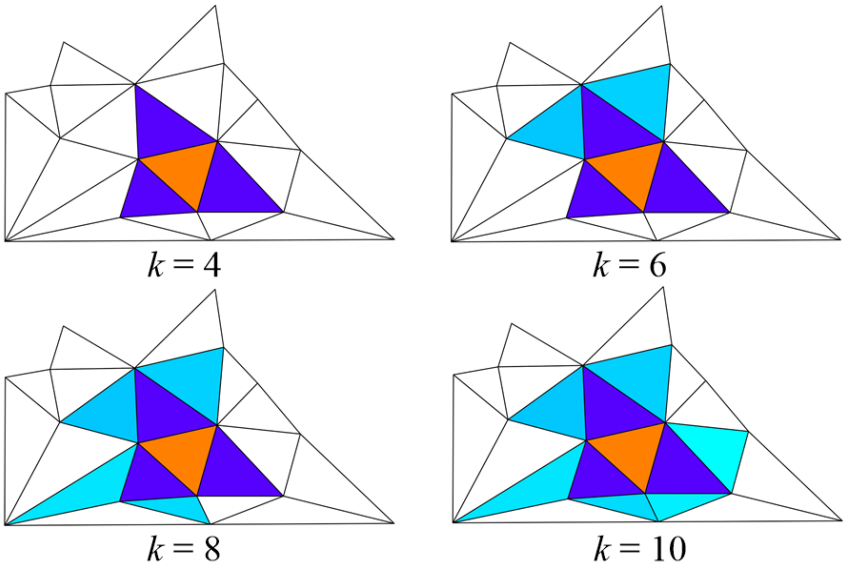

Figure 4. The visualization of convolution receptive fields of different kernel size k. The orange parts are the center faces, which together with the blue faces constitutes the convolutional receptive fields of the center faces.

Point 7: The architecture of this model is still not clear. What are the components of this model and how this model is trained?

Response 7: We apologize for the confusion caused to you. We have added a figure to illustrate the components of the model. On the far right of the figure is the descriptions of the components. In addition, we have added some instructions in section 3.6 (Network architecture) and Section 4.1 (Implementation details) to help readers understand how we train this model. The core of our method has three parts: Face feature (blue), Face convolution (yellow) and Face pooling (green). With these, we can easily generalized 2D convolution networks to triangular meshes. **(p. 8)**

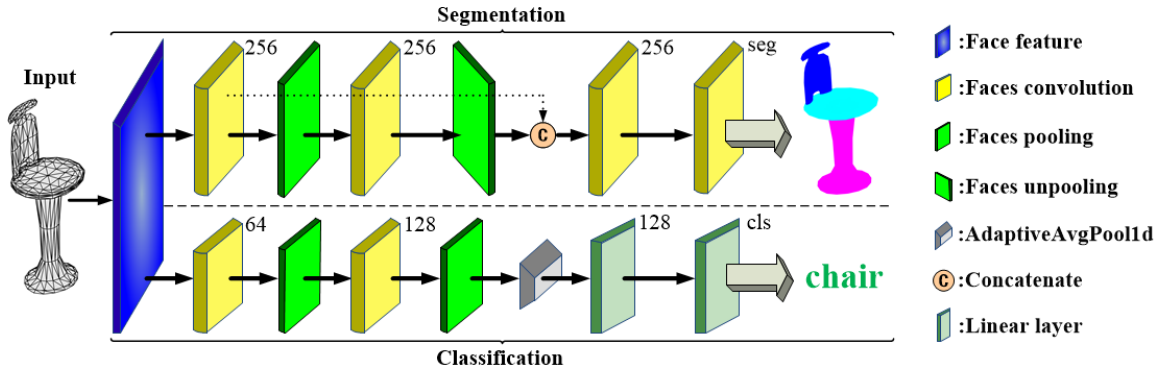

Figure 6. The architecture of network for segmentation and classification, where the top of the block is the number of output channels for the features.