Reaction Report for "O-CNN: Octree-based Convolutional Neural Networks"

Nishant Raj

What I like about this paper?

Since octree-based networks optimally sub-divide the spaces into octants only where shape information is present and use sampled average normal vectors in these octants to further convolve over them, we see that issues pertaining to both memory requirements and computational overheads in dense volumetric networks get solved without compromising on the performance metrics and allowing to work with higher resolutions. Also, since this is a volumetric approach, it overcomes the limitation of obstructions and optimal view selection in view-based approaches. One major good thing about the approach was that since it used continuous arrays at each depth, computations are easy to parallelize in a GPU. The label array used in paper helped to map shape correspondences at different depths. Also, I specifically liked the use of hashing for neighbor discovery which helps a lot in speeding up GPU computations. The overall results are very promising i.e., they are at par to the existing benchmarks. These clearly demonstrate the effectiveness of O-CNN Networks in classification, segmentation, and shape-retrieval tasks.

What I do not like about this paper?

Though the paper opens new dimensions for effectively seeing how operations in volume-based CNN can be optimized, there is little theoretical proof as to why the idea of convolving in an octree lattice structure in regions where object is present is helpful. There is no justification as to why the cubic geometry was selected and if other geometries could yield better performances. Also, authors have demonstrated the results only on ModelNet40 datasets. Though the architecture design seems robust enough, it would be interesting to validate the performances on other datasets which are bigger in size as well for classification tasks. In terms of implementation details, in the orientation pooling part, the authors chose to proceed with voting approach, but fine tuning last two layers might have been more helpful given these are related tasks and could have led to an increase in performance.

Future Directions:

One of the experimentation areas that I would like to try is using fine-tuning approach in orientation pooling part of the paper rather than the max-voting approach. Apart from this, I feel that geometry change in shape has been given very little considerations which authors have also mentioned in the end. An adaptive nature to geometry would ensure that we are able to get best approximation of the given contours of the object. They have improved on this in their subsequent work in Adaptive O-CNN, however there as well, they have only considered planar patches. The improvement in performance there with planar patches certainly helped in understanding that adapting to shape structure is helpful and suggesting trying other forms of geometric patch structures as well. It would also be interesting to see how the O-CNN performs and adapts in other tasks like shape completion, scene reconstruction with little data points apart from the three tasks mentioned in paper i.e., classification, segmentation, and shape retrieval. Finally, as described in the second section, we do not know the theoretical justification of using cubic lattices (though it is more intuitive). It would be interesting to explore theoretical aspect and then try other lattices as well and see if we can get an improvement in optimization in terms of computation and memory requirements.