

1. When talking about 3D reconstruction, we said that a single image does not have enough information to reconstruct the 3D scene. How then is the PTN model able to reconstruct objects from a single image?
2. Is the trained PTN model *guaranteed* to produce a 3D shape that is consistent with the input image? If not, can you do something at inference to get this guarantee?
3. Does the PTN model or training pipeline use any insight from the geometry of image formation?
4. During training, the model is provided with multiple views of the same object. This is also the typical setup for stereo. Can you think of ways of using off-the-shelf stereo techniques to improve the model?
5. PTN represents the final object shape as a grid of “voxels”. What issues do you see with this representation?