

# PerspectiveNet: A Scene-consistent Image Generator for New View Synthesis in Real Indoor Environments

Kobee Raveendran

February 9, 2020

## 1 Summary

This paper tackles the problem of novel view synthesis. Given a set of images of an indoor location taken from unique viewpoints, the authors' method attempts to generate the view from another, new viewpoint. The authors describe their primary contribution as a new take on the render-inpaint approach to novel view synthesis that treats the reference view and generated view as a calibration task. Their method uses the regular geometric properties of indoor environments to generate the new views.

## 2 Good points

Something I found strong about this paper was the complexity of the problem the authors are attempting to solve. Previous methods were less flexible and could usually only perform well on extremely constrained tasks (such as generating viewpoints of specific objects). The proposed method can effectively extend to generate views of the entire contents of a room, which is impressive given that rooms are not "well-defined" and as such are quite unconstrained in nature. Additionally, the input is relatively simple and does not require extensive pre-processing or tuning to be fed into the method.

## 3 Weak points

Overall this paper seemed well-rounded, with abundant quantitative and qualitative metrics shown, promising results, and addressing a complex task. The improvements shown in the quantitative results are very minimal (typically only outperforming previous methods by less than one unit of measurement). However, this could be offset by their impressive performance in their qualitative evaluation. Additionally, another potential weakness could be the limited data set sample; only two data sets were used. This could possibly be a fault of the niche problem space and subsequent lack of established data sets, however I would imagine that it would be fairly easy to generate data (especially since it is mostly just interior room pictures that could be taken with handheld cameras).

## 4 Questions

To address problems occurring during in-painting that can lead to vastly differing room aesthetics, the authors include additional regularizers to keep the scene more or less attached to the ground truth. This regularizer has two terms, both of which control the "difference" between the ground truth images and the generated view. This leads me to wonder why there should be much of a difference at all; if the scene in the room is meant to be consistent across the viewpoints, why is this difference not being completely minimized (thus only accounting for differences in pixel values due to perspective change, rather than introduction of things like new objects)?

## 5 Ideas

The authors mention a certain module of their approach that generates a point cloud from solely image data. This is interesting to me because I was unaware that this was possible; if this 'point cloud' is anything like the point clouds LIDAR sensors produce, and that depth and distance can be accurately portrayed from image data alone, it would be interesting to see how these image-only point clouds could be used in applications such as autonomous driving.