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

Abraham Jose

1 Summary

PerspectiveNet is a novel work in scene and geometric consistent image-generator for a new view synthesis, for a given set of reference RGBD images. The stages consist of a RGBD autoencoder for inpainting task on a large dataset of partial renders and a scene-level optimizer which uses another autoencoder, altered by learnable feature residuals. The introduced losses like reprojection consistency loss, style consistency loss and scene-level optimization helps to achieve state-of-the art result quantitatively and qualitatively in SceneNet and ScanNet dataset.

2 Good points

The proposed model outperforms the considered baselines on all depth metrics. They shows that scene-consistent optimizer brings out better image realism using their ablation study using LPIPS metrics. The performance is qualitatively and quantitatively compared with other state of the art techniques which is commendable also they provide a comprehensive summary on why the other techniques why they lack performance w.r.t. each metrics. Using deep latent coding of depth images from the auto-encoder overcomes the need of complex regularizers.

3 Weak points

The auto-encoder model(ϕ) pretrained on imageNet is taking 7 days using single GPU to train the model in SceneNet and ScanNet because the training is limited to 4 batches at once. Also the inference time for PrespectiveNet, regularizers and scene level optimizations is not mentioned in paper. However, it is clear that it is not possible to create real-time results which is required in many applications which is a drawback.

4 Questions

How well PrespectiveNet performs if the test contains few content(less than 50% for ScanNet/ 40% for SceneNet) from the 4 training image content? And how well the differentiable point tracer technique could account for occlusion those such cases?

5 Ideas

The bundle adjustment that optimizes all the different views can be introduced to SLAM techniques like ORB-SLAM and RSLAM where the loop closing and bundle adjustments are required which will increase the performance.