EdgeNet: Semantic Scene Completion from a Single RGB-D Image

Aloisio Dourado, Teofilo Emídio de Campos, Hansung Kim, Adrian Hilton
aloisio.dourado.bh@gmail.com, t.decampos@st-annes.oxon.org, H.Kim@soton.ac.uk, a.hilton@surrey.ac.uk

Given an RGB-D image, the goal of semantic scene completion is to infer a complete 3D occupancy grid with associated semantic labels. Previous works completely neglect the RGB channels from the input data or require a complex two-step training process to merge RGB and depth data.

Our Edge-Net representation encodes colour information in 3D space using edge detection and flipped truncated signed distance (F-TSDF), which improves semantic completion scores especially in hard-to-detect classes, with an end-to-end 3D deep neural network.

Our deep CNN is trained on SUNCG and fine-tuned on NYUDv2.

Using edges and F-TSDF, we address the data sparsity problem faced by previous solutions that tried to explore the RGB components of the RGB-D data. Our solution is an end-to-end network architecture that may be trained as a whole and achieves state-of-the-art results.

DATASETS
SUNCG - Synthetic Scenes
NYUDv2 – Indoor RGB-D Scenes

SUMMARY OF THE CONTRIBUTIONS
• A new end-to-end network architecture
• A new strategy to encode information obtained from RGB
• Improvement over the state-of-the-art result on SUNCG
• An efficient and lightweight training pipeline for the task