Joint Inpainting of RGB and Depth Images by Generative Adversarial Network with a Late Fusion approach

Ryo Fujii, Ryo Hachiuma, Hideo Saito
Keio University, Japan
{ryo.fujii0112, ryo-hachiuma, hs}@keio.jp

Abstract

Task: Simultaneous RGB and depth image inpainting

Input: RGB and depth image with missing regions
Output: inpainted RGB and depth image

We first propose a deep learning network that jointly inpaint RGB and depth image while leveraging each other's information.

Goal:
Filling in missing regions of both RGB and depth images with plausible textures and geometries.

Contribution:
The output features of RGB and depth encoder are added and used as the input of fusion part. This late fusion enables each decoder to utilize each feature as complementary information.

Method

Network Architecture:
We expanded the completion network proposed by Iizuka et. al.[5]

Completion Network
• RGB encoder-decoder
• Depth encoder-decoder
• Fusion Part

The extracted feature of RGB and depth are fused and we employed Residual dilated convolutional layers (yellow) in Fusion part.

Discriminator Network
Input: four channels, RGB-D image.
• Global discriminator: Judges the consistency of scene
• Local discriminator: Assesses the quality of small completed area

Dataset:
SceneNet RGB-D dataset [6]
Rendered RGB-D images from over 15K trajectories in synthetic layouts.

Training procedure:
• Training images: About 2 million images resized to 160 x 160
• Mask: 1/8 to 1/4 of original size
• Batch size: 96 images
• Iteration: 75000 iterations
• Training time: About 2 days by two Nvidia Quadro GV100 and P6000 GPU
• Optimizer: Adam

Result

Quantitative evaluation:

Input RGB or Depth

Generated Result

Ground truth

• Late fusion approach made the edge clear of each restored region clearly.
• The depth completion sometimes fails as shown in third column of above figure.
  ➢ In the most common failure case, the similar texture appears.

Qualitative evaluation:

Future work:
• Quantitative evaluation
  ➢ Comparison our baseline model with early fusion approach
• Examination of the input
  ➢ Take advantage of normal map instead of depth map.
• Application
  ➢ Integrate this model into virtual reality application with HMD
  Filling holes because foreground object occlusion due to the viewpoint changes caused by the user's head movements.