Single Noisy Image Depth Estimation with Multi-Scale Feature Fusion Module

Yuxiao Chen  
yuxiaoc@stanford.edu

Qingxi Meng  
qingxi@stanford.edu

Introduction

- Monocular depth estimation is ill-posed due to the lack of information [1]
- Stereo information more expensive and unsuitable for deployment [2]
- Gaussian noise stereo image reconstruction
  - Explore median and bilateral denoise filter
  - Explore multi-scale feature fusion module (MFF) [3]

Model Architecture

- Denoise: Median / Bilateral
- Depth Estimation: encoder (E), decoder (D), multi-scale feature fusion module (MFF), and refinement module (R)

Experiments

- Trained on 8000 uniformly sampled rgb-depth pairs
  - Add Gaussian noise $\sigma=0.1$ to the RGB image input
  - Use depth loss, normal loss, and gradient loss
- Evaluate with 150 unseen rgb-depth image pairs

Discussion

<table>
<thead>
<tr>
<th>Model Trained without Denoise</th>
<th>Model Trained without Denoise</th>
<th>Model Trained with Median Filter</th>
<th>Model Trained with Bilateral Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>without MFF</td>
<td>0.660</td>
<td>0.594</td>
<td>0.697</td>
</tr>
<tr>
<td>with MFF</td>
<td>0.660</td>
<td>0.594</td>
<td>0.697</td>
</tr>
</tbody>
</table>

Challenges / Limitations

- Denoise filter can smooth out edges, impede the depth-estimation performance
- Real noise have unknown distribution

Future Work

- Explore other denoise models (Non-Local Mean, DnCNN)
- Explore other depth estimation methods (self-supervised)

Summary

- Training with MFF yields lower loss than without MFF
- Denoise slightly improves depth estimation

Reference