Semantic-Guided Inpainting Network for Complex Urban Scenes Manipulation

Pierfrancesco Ardino, Yahui Liu, Elisa Ricci, Bruno Lepri, Marco De Nadai
The art of transforming an image to convey what you want

Image Manipulation

- Object Insertion
- Object removal or image inpainting

+ "car" label →

+ none →
Challenges

- Mixed scenes contains objects and background of different classes
Challenges

- Missing part with different semantics

Image manipulation in complex urban scenarios

- Object insertion
- Image reconstruction

1. Remove objects via input shape + none
2. Remove objects / area inpainting + none
3. Insert objects via random sampling + “car” label
4. Insert objects via input shape +
SGI-Net: Architecture

Incomplete image ($X_{\text{BLANKED}}$)

Segmentation ($S_{\text{BLANKED}}$)

Instance sampled code ($Z$) and class ($c$)

Completed image ($X_{\text{FILLED}}$)

Completed segmentation ($S_{\text{FILLED}}$)

Generated instance ($\hat{m}_s$)

Instance inserted

$E_{\text{im}}$

$E_{\text{se}}$

ResBlocks

$G$

$D_g$

$D_s$

real / fake

real / fake
SGI-Net: Decoder Block

Our Decoder Block (inside the G)

Conv

Pixelshuffle (2)

Log Softmax

3x3 Conv

SPADE ResBlock

SPADE

ReLU

3x3 Conv

SPADE

ReLU

3x3 Conv

Multi-scale segmentation loss

$s_{k-1}^{FILLED}$

$s_{k}^{FILLED}$

$s_{k+1}^{FILLED}$
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Experimental setting</th>
<th>Cityscapes</th>
<th>Indian Vista</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PSNR↑</td>
<td>FID↓</td>
</tr>
<tr>
<td>Hong et al. [12]</td>
<td>Restore</td>
<td>31.07</td>
<td>7.26</td>
</tr>
<tr>
<td>SFG-Net [41]</td>
<td></td>
<td>31.36</td>
<td>7.97</td>
</tr>
<tr>
<td>RN [50]</td>
<td></td>
<td>32.16</td>
<td>9.64</td>
</tr>
<tr>
<td>Our proposal</td>
<td></td>
<td><strong>32.95</strong></td>
<td><strong>5.08</strong></td>
</tr>
<tr>
<td>Hong et al. [12]</td>
<td>Place</td>
<td>31.08</td>
<td>7.26</td>
</tr>
<tr>
<td>SPG-Net*</td>
<td></td>
<td>31.37</td>
<td>7.96</td>
</tr>
<tr>
<td>RN*</td>
<td></td>
<td>31.74</td>
<td>9.79</td>
</tr>
<tr>
<td>Our proposal</td>
<td></td>
<td><strong>32.96</strong></td>
<td><strong>5.05</strong></td>
</tr>
</tbody>
</table>

**Input image**

- Hong et al. [12]
- SPG-Net
- RN
- Our

**Car**

- Hong et al. [12]
- SPG-Net*
- RN*
- Our
- Our (segmentation)
- Ground truth
Thank You

Github: https://github.com/PierfrancescoArdino/SGINet