Learning Image Inpainting from Incomplete Images using Self-Supervision

Sriram Yenamandra, Ansh Khurana, Rohit Jena, Suyash P. Awate

Computer Science and Engineering Department, Indian Institute of Technology Bombay
Outline

Introduction

Our method

Results
Motivation

- Semantic image inpainting refers to the task of restoring missing parts of a corrupted image using the available data.
- Current state-of-the-art deep learning based image inpainting methods are fully-supervised i.e., require complete images for learning.
- Obtaining large number of complete images is infeasible in many applications like brain tumor removal in MRI images.
- Motivates the need to learn to inpaint images using a dataset having incomplete images.
Contributions

- We propose a self-supervised framework that can learn to inpaint in both semi-supervised and fully unsupervised settings.
- Our method trained only using incomplete images outperforms state-of-the-art learning under full supervision.
- Our method leads to more stable training as it does away with adversarial training and density estimation in higher dimensional spaces.
Introduction

Our method

Results
• The known regions of an incomplete image can be utilized for training our DNN (UNet)
• Self-supervised learning: Introduce holes in the incomplete images and task the DNN to complete the input image
• Higher weight to the training loss on introduced regions as compared to other known regions
The pixels are grouped into three categories: type-A pixels - originally missing in $Y$, type-B pixels - present in $Y$ but removed in $Z$ and type-C - present in $Y$ and left unchanged in $Z$.

The loss is contributed by type-B and type-C pixels. The parameter $\alpha$ used to weigh the contributions; $\alpha = 0.75$ found to work best.
Experiments

- We evaluate our method on CelebA face images dataset
- Create incomplete images by introducing corruption in smooth regions of randomly generated shapes and sizes
- We train our method for different levels of supervision by varying the fraction of complete images ($\gamma$)
- We use structural similarity (SSIM) and relative root-mean-squared-error (RRMSE) for quantitative evaluation
Quantitative Results

- Our method’s performance is compared to fully-supervised inpainting methods like PIC, Shift-Net and VAEAC.

Table 1: Results of All Methods Trained on the Entire CelebA Dataset

<table>
<thead>
<tr>
<th>Method</th>
<th>Data, Training Mode</th>
<th>SSIM mean(std.dev.)</th>
<th>RRMSE mean(std.dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>$\gamma = 1$, Fully Supervised</td>
<td>0.938 (0.019)</td>
<td>0.055 (0.018)</td>
</tr>
<tr>
<td>Ours</td>
<td>$\gamma = 0$, Unsupervised</td>
<td>0.936 (0.019)</td>
<td>0.055 (0.020)</td>
</tr>
<tr>
<td>VAEAC</td>
<td>$\gamma = 1$, Fully Supervised</td>
<td>0.913 (0.024)</td>
<td>0.129 (0.043)</td>
</tr>
<tr>
<td>PIC</td>
<td>$\gamma = 1$, Fully Supervised</td>
<td>0.907 (0.021)</td>
<td>0.085 (0.027)</td>
</tr>
<tr>
<td>SNet</td>
<td>$\gamma = 1$, Fully Supervised</td>
<td>0.891 (0.022)</td>
<td>0.095 (0.029)</td>
</tr>
<tr>
<td>DIP</td>
<td>Not Applicable</td>
<td>0.885 (0.021)</td>
<td>0.106 (0.031)</td>
</tr>
</tbody>
</table>
Quantitative Results

Results on varying level of supervision

![Graph showing SSIM and RRMSE values for different levels of supervision with various methods compared: PIC, Ours, DIP, VAEAC, SNet.]
Qualitative Results

Varying the level of supervision (percentage of complete images)

<table>
<thead>
<tr>
<th>Ours (γ = 0.0)</th>
<th>Ours</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>VAEAC</td>
<td>PIC</td>
</tr>
<tr>
<td>DIP</td>
<td>SNet</td>
<td>0.1</td>
</tr>
<tr>
<td>Input</td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>Level of supervision (γ)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Qualitative - Different Masks

Our model trained unsupervisedly generalizes well across different mask distributions
Qualitative - Application
Summarizing our results

- Our method can leverage incomplete images and produce high quality inpaintings
- The performance doesn’t drop on decreasing the level of supervision
- Our method outperforms the state-of-the-art inpainting algorithms both quantitatively and qualitatively
- It generalizes well across different mask distributions
Thank You!