# CS663: Digital Image Processing - Homework 5

Harsh | Pranav | Swayam

November 6, 2024

## Homework 5 - Question 3

### Referenced Paper

**Title** : Image Inpainting for Irregular Holes Using Partial Convolutions
**Authors** : Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, Bryan Catanzaro
**Venue** : ECCV 2018
**Link** : Image Inpainting for Irregular Holes Using Partial Convolutions

### Problem Statement

The paper addresses the problem of image inpainting specifically for irregular holes. Traditional inpainting techniques struggle with irregular hole patterns and often rely on post-processing, leading to artifacts in the generated images, such as blurriness and color inconsistencies. The method presented by Liu et al. improves the robustness and quality of inpainting by using partial convolutional layers that only consider valid (non-masked) pixels during the convolutional process. This approach ensures the inpainting model can effectively handle various hole sizes and shapes, without relying on post-processing.

### Key Concepts

1. **Partial Convolutions**: Unlike standard convolutional layers that use fixed values for missing pixels, partial convolutions condition their output based solely on valid (non-masked) pixels.
2. **Mask Update Mechanism**: The model includes an automatic mask update step, which adjusts the mask during the forward pass based on which pixels were successfully processed by the convolution operation.

### Cost Function

The optimization process involves several loss functions aimed at ensuring both reconstruction accuracy and smooth transitions around the inpainted areas. The key components of the loss function include:

**Per-Pixel Reconstruction Loss**

For hole pixels:
$$L_{\text{hole}} = ||(1 - M) \odot (I_{\text{out}} - I_{\text{gt}})||_1$$

For valid pixels:
$$L_{\text{valid}} = ||M \odot (I_{\text{out}} - I_{\text{gt}})||_1$$

Here:

- $I_{\text{out}}$: The output image produced by the network.

- $I_{\text{gt}}$: The ground truth image.

- $M$: The binary mask indicating hole locations (0 for holes, 1 for valid pixels).

- $\odot$: Element-wise multiplication.

- $|| \cdot ||_1$: L1 norm, measuring absolute differences.

**Perceptual Loss**

This loss assesses how well the output image captures high-level features compared to the ground truth by using a pretrained VGG-16 network:

$$L_{\text{perceptual}} = \sum_{n=0}^{N-1} ||\Psi_n(I_{\text{out}}) - \Psi_n(I_{\text{gt}})||1 + \sum n = 0^{N-1}||\Psi_n(I_{\text{comp}}) - \Psi_n(I_{\text{gt}})||_1$$

where $\Psi_n$ represents activation maps from selected layers of VGG-16.

*VGG-16 is a convolutional neural network architecture developed by the Visual Geometry Group at the University of Oxford. It is renowned for its simplicity and effectiveness in image classification tasks. The architecture consists of 16 layers with learnable parameters, including 13 convolutional layers and 3 fully connected layers.*

**Style Loss**

This term measures the similarity of style between images using Gram matrices derived from feature maps:

$$L_{\text{style}} = ||G(I_{\text{out}}) - G(I_{\text{gt}})||_1$$

where $G(\cdot)$ computes the Gram matrix of features.

## Overall Loss Function

The total loss function combines these components to guide the training process:

$$L = L_{\text{hole}} + L_{\text{valid}} + \lambda_1 L_{\text{perceptual}} + \lambda_2 L_{\text{style}}$$

where $\lambda_1$ and $\lambda_2$ are hyperparameters that control the weight of perceptual and style losses relative to reconstruction losses.