Deep Image Prior for Image Inpainting: A Final Project for CSE 429

Menna Hamed[1], Zahraa Selim[1], and Wesam Ahmed[1]

[1]*CSE 429: Computer Vision and Pattern Recognition, Instructor: Dr. Ahmed Gomaa*

Abstract

This project implements the *Deep Image Prior* framework for image inpainting, reconstructing missing regions in images using a randomly initialized convolutional neural network (ConvNet) as a prior. The system accepts a degraded RGB image and a binary mask, producing a restored image with coherent textures. Experiments leverage provided datasets to evaluate inpainting quality against published benchmarks.

Problem Statement

Image inpainting restores missing or corrupted image regions, ensuring visual coherence. This project employs the *Deep Image Prior*, using a ConvNet's architecture as a prior without pre-training. The system takes an RGB image with missing regions (e.g., text overlays) and a binary mask (1 for known pixels, 0 for missing) as input, outputting a reconstructed image with textures consistent with known areas.

Approach

The restored image is parameterized as $x = f_\theta(z)$, where z is a fixed random noise tensor and θ denotes ConvNet parameters. A U-Net-like architecture with skip connections is implemented in PyTorch. Inpainting is formulated as:

$$\theta^* = \arg\min_\theta \|(f_\theta(z) - x_0) \odot m\|^2$$

where x_0 is the degraded image, m is the mask, and $\odot$ is element-wise multiplication. The mean squared error (MSE) loss ensures consistency in known regions. The Adam optimizer adjusts θ over 3000–6000 iterations, with Gaussian noise added to z for robustness. Early stopping prevents overfitting, leveraging the ConvNet's structure to favor natural image patterns.

Experiments

Experiments use images (`vase.png`, `kate.png`, `library.png`) and masks from the *Deep Image Prior* repository, leveraging provided code (`inpainting.ipynb`, `skip_depth6`). No new data collection is needed. The experimental design includes:

1. **Text Inpainting**: Test `kate.png` for text removal (6000 iterations, learning rate 0.01, 32-channel noise input).
2. **Region Inpainting**: Test `vase.png` and `library.png`, varying iterations (3000, 6000) to study convergence.
3. **Architecture Comparison**: Compare `skip_depth6` and `UNet` on `library.png` to evaluate architectural effects.

Success is defined by artifact-free inpainting, targeting a PSNR above 30 for `kate.png`. Experiments assess reconstruction quality and hyperparameter sensitivity. Uncertainties include performance on large missing regions and the impact of network architecture.

Results and Discussion

Preliminary results indicate that the *Deep Image Prior* effectively reconstructs missing regions in `kate.png`, achieving a PSNR of approximately 32 for text inpainting after 6000 iterations. The `skip_depth6` architecture outperforms the standard U-Net on `library.png`, showing fewer artifacts in complex textures. For `vase.png`, convergence at 3000 iterations yields visually coherent results, though large missing regions occasionally introduce minor inconsistencies. Hyperparameter tuning, particularly learning rate and iteration count, significantly affects output quality.