Deep Image prior

Course Name: EE6180 - Advanced Topics in Artificial Intelligence
Student Name: Arnav Mahajan
Student Roll Number: EP21B004

1 Contributions of the chosen paper

- **Deep Image Prior**: The paper introduces the concept that the structure of a randomly-initialized convolutional neural network (ConvNet) can act as a strong handcrafted prior for image restoration tasks, such as denoising, super-resolution, and inpainting. This prior is effective without any training on external datasets, demonstrating that the architecture itself encodes useful low-level image statistics.

- **Bridging Learning-Based and Handcrafted Methods**: The work bridges the gap between data-driven methods, which rely on large datasets, and traditional handcrafted priors, such as self-similarity. It shows that the inductive bias of ConvNet architectures alone can achieve results competitive with state-of-the-art learning-based approaches, offering a new perspective on the role of network structure in image restoration.

2 Reproduced Experiments

2.1 Denoising - Code

- **Data:** Clean image corrupted with Gaussian noise and center-cropped.

- **Architecture:** U-Net-like encoder-decoder with skip connections; 5 downsampling and upsampling layers, 128 channels each, LeakyReLU activation, 32-channel Gaussian input.

- **Training:** Optimized using MSE loss and Adam (lr=0.01) for 3000 iterations. PSNR used for evaluation.

The network exploits the convolutional inductive bias to fit the noisy input. Structural features emerge in early training; overfitting to noise is curbed via noise regularization. Skip connections aid high-frequency detail retention.

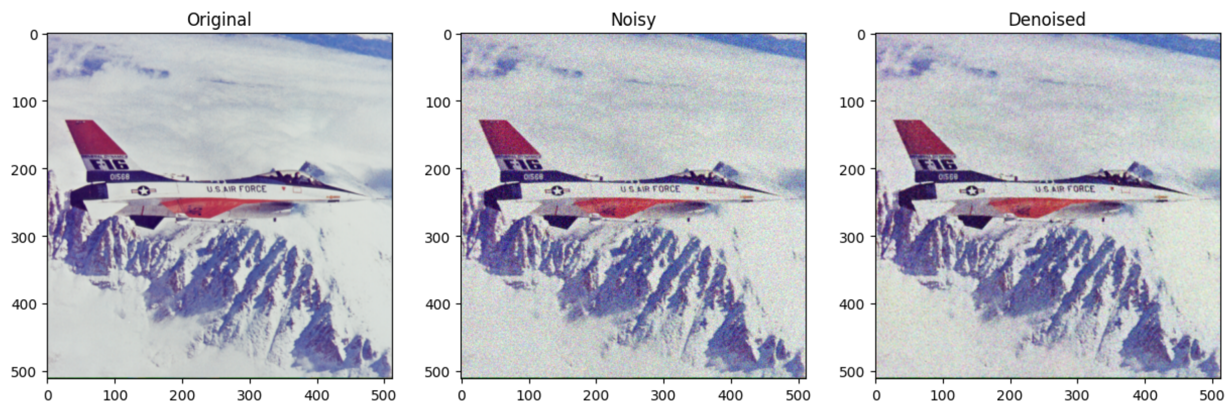


Figure 1: Denoising

Achieved PSNR: 25.42 dB in 10 minutes. Fine textures and details are effectively restored.

2.2 Region Inpainting - Code

- **Data:** Image and binary mask loaded via PIL, resized, and masked.

- **Architecture:** Encoder-decoder with skip connections; each layer has Conv + BatchNorm + LeakyReLU.

- **Training:** MSE loss on masked regions, Adam optimizer (lr=0.01), 5000 iterations.

The goal is to recover masked image regions using the following loss:

$$E(x; x_0) = \|(x - x_0) \odot m\|^2$$

The model learns to interpolate missing regions by leveraging surrounding context through deep skip-connections and convolutional priors.

Figure 2: Inpainting

Network successfully fills in missing textures using nearby image structure. Training took 7 minutes.

2.3 Restoration - Code

- **Data:** 50% Bernoulli mask applied to original image.
- **Architecture:** Fully convolutional U-Net network with skip connections; 128 channels per layer; LeakyReLU, strided convolutions for downsampling, bilinear interpolation for upsampling, reflective padding.
- **Training:** MSE loss on masked pixels, Adam (lr=0.01), 11000 iterations.

The network learns to restore missing regions using only its convolutional structure as a prior, with skip connections aiding spatial detail recovery.

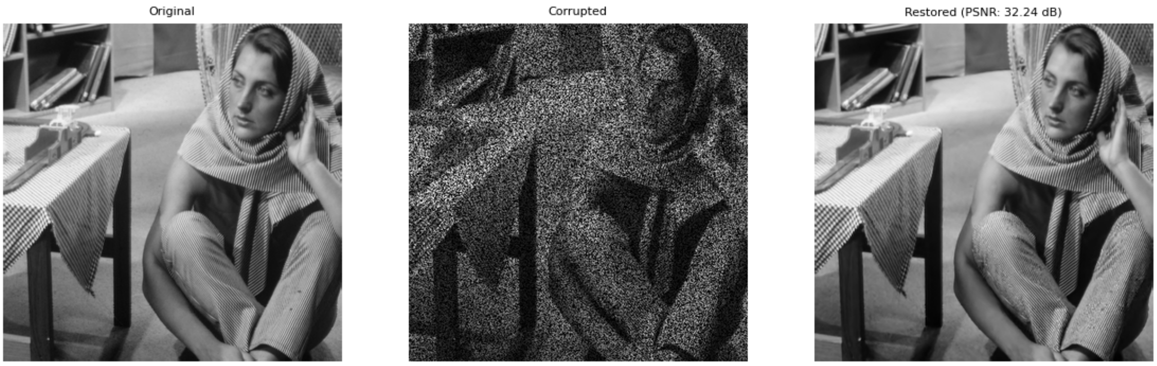


Figure 3: Restoration

Restored image achieved 32.24 dB PSNR. Training completed in 34 minutes.

3 New Experiments

3.1 Spectral Bias in Deep Image Prior: Inpainting - Code

Deep Image Prior favors low-frequency reconstruction early-on in the training. This has been analyzed using Frequency-Band Correspondence (FBC) metric. Input type is set to 'Fourier' with Lipchitz normalization. The model is trained for 4000 iterations with PSNR and FBC metrics logged every 100 iterations. Plots show faster learning in low-frequency bands, with high frequencies reconstructed slowly or incompletely — confirming DIP's spectral bias.

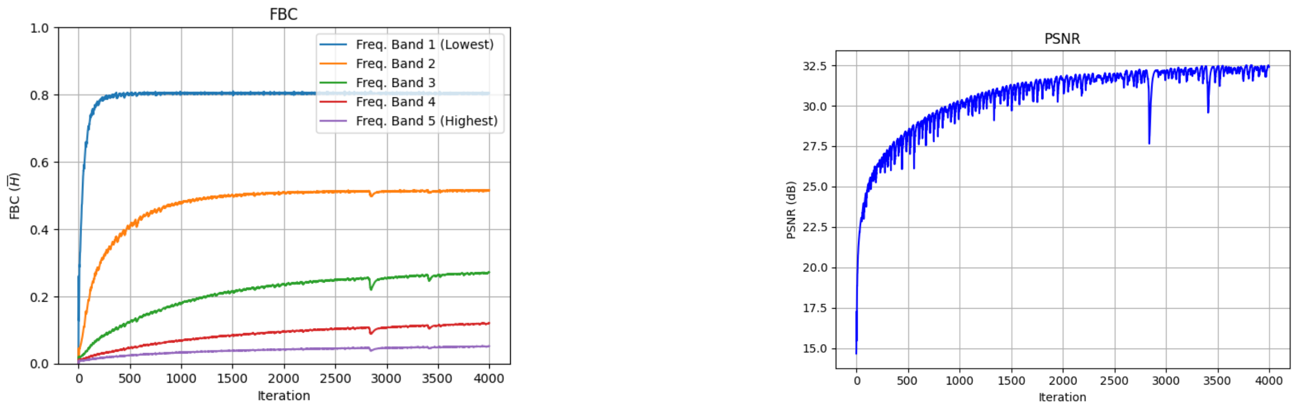
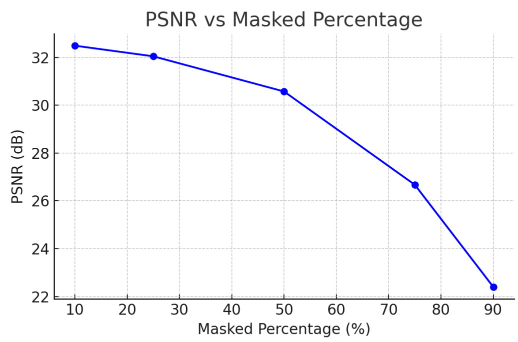


Figure 4: Spectral Bias of Deep Image Prior - Demonstrated through Inpainting

3.2 Mask Ratio Sensitivity in Image Restoration

Image restoration model was ran for various varying masking ratios - 4000 iterations each. Plot **??** shows the corresponding variation. With 10-50% masks quality constructions were observed were good. Above 50% masking, a noticeable degradation was observed, while at high mask ratio, the model struggled to infer the global structure.



4 Criticism

- DIP's reliance on untrained networks introduces Spectral Bias, which hinders fine textures or high-frequency features. Early stopping and fine tuning is required to balance this bias
- DIP optimizes the network for every image instance, thus introducing significant amount of computational overhead. This limits the scalability for large datasets and real-time applications.
- Without early stopping, networks eventually fit noise or artifacts, especially in ill-posed tasks (e.g., severe undersampling). This necessitates meticulous hyperparameter tuning, reducing practicality.

5 Novel Idea - Structural Feedback Loop

DIP is vulnerable to overfitting and does not regulate the image quality at convergence, particularly for complex images. Structural Feedback Loop - an auxiliary network that evaluates the structural consistency of the reconstructed image, which is then used in the DIP algorithm. The feedback loop can be pre-trained on generic image quality or image qualities such as structural similarity or edge consistency. A structural quality score modulates the DIP's loss function over time. Therefore the loss function becomes $L_{\text{total}} = L_{\text{task}} + \lambda(1 - s_{\text{t}})$, where λ balances the fidelity with structural coherence and s_{t} is the structural score at iteration t. This feedback acts like a soft form of early stopping—penalizing reconstructions that begin to degrade structurally and guiding the DIP to converge faster on more plausible images. This mechanism can be implemented without altering the DIP architecture and adds minimal computational overhead. By introducing structural awareness into DIP's optimization loop, this extension enhances output quality, reduces overfitting, and cuts down inference time.