Image Super-Resolution Using Deep Learning

Laxmi Sreenivas IMT2020510
Rithvik Ramasani IMT2020543

Problem Statement

Problem : Given a low-resolution image, we need to develop a deep learning model to create a high-resolution version by inferring missing image details.

Challenges

- **Ill-posed problem:** There can be multiple high-resolution images that could have resulted in the same low-resolution image. The model needs to learn the best possible reconstruction.
- **Preserving detail:** The model needs to not only increase resolution but also recover lost details and textures from the low-resolution image.
- **Avoiding artifacts:** The reconstruction process should not introduce unwanted artifacts like noise or unnatural sharpening.

Literature Review & Motivation (0)

Convention models for SR tasks :

- **Convolutional Neural Networks(CNN) based models:**
  - Computationally efficient.
  - Can learn well from smaller datasets due to their inductive bias (ability to generalize from limited data).
  - However, CNNs struggle to capture long-range dependencies between different parts of the image.
- **Vision Transformer (VT) based models:**
  - Excel at capturing long-range dependencies due to their self-attention mechanism.
  - However, require a large amount of training data and computational resources, limiting their use in scenarios with limited data and GPUs.
  - VTs lack the inductive bias of CNNs.

Literature Review & Motivation (1)

- **Wavelet based models:**
    - Leverage the power of wavelet transforms to decompose images into approximation (coarse) and detail (fine) information.
    - *Advantages:*
        - Wavelet decomposition simplifies mapping between low-resolution and high-resolution images.
        - Invertible wavelet transform ensures no information loss during downsampling
    - *SR with Wavelets:*
        - Models like DWSR use CNNs to predict detail coefficients in wavelet space for HR image reconstruction.
        - WaveletSRNet focuses on faces, predicting wavelet coefficients for reconstruction.
        - MWCNN replaces pooling with wavelets in CNNs, preserving information for better long-range dependency capture.

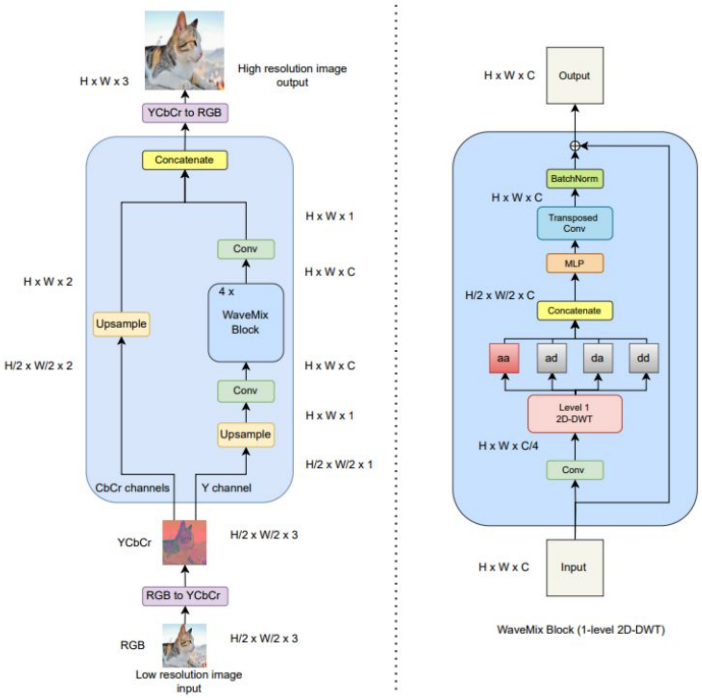
WaveMixSR (0)

- Wavelet based model, that combines the strength of both CNN and VT based models.
- Utilizes spatial token-mixing using a 2D discrete wavelet transform (DWT) for efficient communication between different image regions.
- Usage of 2D DWT, makes it computationally efficient and requiring less training data compared to pure VT models.

- **Architecture (Dual-Path Processing):**
  - Y-channel (luminance): Prioritized for detail reconstruction due to higher information content.
  - CbCr channels (chrominance): Upsampled directly.

WaveMixSR (1)

- **WaveMix Blocks (Y-channel path):**
  - Core component for detail reconstruction and sharpening.
  - Convolution for learnable feature extraction.
  - 2D-discrete wavelet transform (DWT)
  - MLP for channel-wise feature mixing.
  - Transposed convolution to restore spatial resolution.
  - Residual connection for improved gradient flow.
- **Upsampling:**
  - Efficient non-parametric methods (bilinear or bicubic interpolation).



Research Gaps

- **Exploration of Variations in Wavelet Transform:** The study uses 2D-discrete wavelet transform (DWT) for spatial token-mixing. Investigating the effects of using different wavelet families or modifying the DWT implementation could potentially improve performance or efficiency.

- **Incorporating External Knowledge:** The paper focuses on the WaveMixSR architecture itself. How WaveMixSR might benefit from incorporating external knowledge sources, such as image priors or pre-trained, is an unexplored area.

- **Incorporating Adversarial Training:** Adversarial Training has been shown to improve the visual quality of super resolved images compared to normal training. WaveMixSR with adversarial training is an unexplored research zone.

Our Contribution & Experiments (0)

- **Wavelet Family Selection**
  - **Objective:** To determine the best wavelet family for WaveMixSR.
  - **Exploration:** Examined three wavelet families—Haar, DB1, and Symlet.
  - **Evaluation Criteria:** Measured performance using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), focusing on noise levels and image border artifacts.
  - **Outcome:** Haar wavelet provided the best results, with high PSNR and SSIM, and minimal noise. This led to its selection for the WaveMixSR.

Our Contribution & Experiments (1)

- **Pre-Trained Network Selection**
  - **Objective:** To enhance feature extraction capabilities for WaveMixSR.
  - **Pre-Trained Networks:** Considered VGG11 and ResNet18, selecting specific layers for feature extraction.
  - **ResNet18 Selection:** Outperformed VGG11, offering more accurate image representation and residual learning benefits.
  - **Selected Layers:**
    - conv1, bn1, relu, and maxpool: Early feature extraction and spatial downsampling.
    - layer1 and layer2: Residual blocks for complex feature learning.
  - **Outcome:** Integration of ResNet18 provided a strong foundation for feature extraction, leading to more accurate super-resolution results.

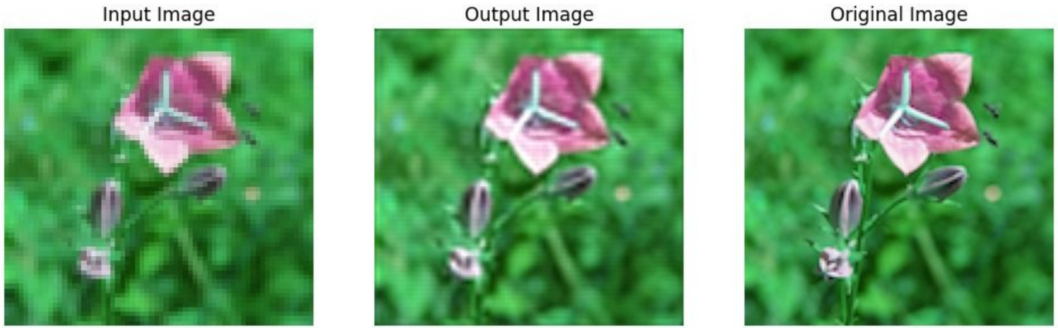
Our Contribution & Experiments (2)

- **Adversarial Training Integration**
  - **Objective:** To create more realistic high-resolution images through adversarial training.
  - **GAN Framework:** Introduced a generator (WaveMixSR with ResNet18 backbone) and a discriminator for adversarial training.
  - **Discriminator Design:**
    - Combination of convolutional layers, batch normalization, and fully connected layers.
    - Dropout applied to prevent overfitting.
    - Role: Distinguish real images from generated images, challenging the generator to improve.
  - **Generator Enhancement:** The WaveMixSR generator, backed by ResNet18, was improved through adversarial training to produce more realistic outputs.
  - **Outcome:** Adversarial training led to significant improvements in image quality, with the generator producing more realistic high-resolution images. The setup enhanced both perceptual quality and computational efficiency, demonstrating the effectiveness of integrating GAN-based approaches into super-resolution tasks.

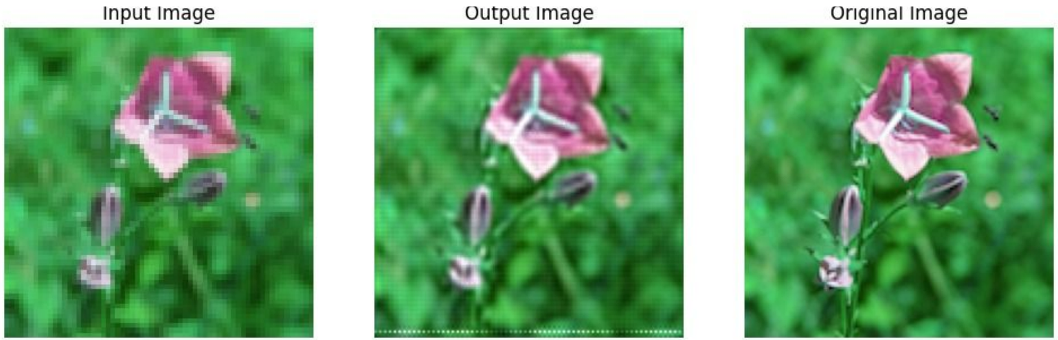
Results and Discussion (0)

- **Results**
  - **PSNR:**
    - Haar Wavelet: 29.1
    - DB1 Wavelet: 28.45
    - Symlet Wavelet: 28.7
  - **SSIM:**
    - Haar Wavelet: 0.87
    - DB1 Wavelet: 0.79
    - Symlet Wavelet: 0.81

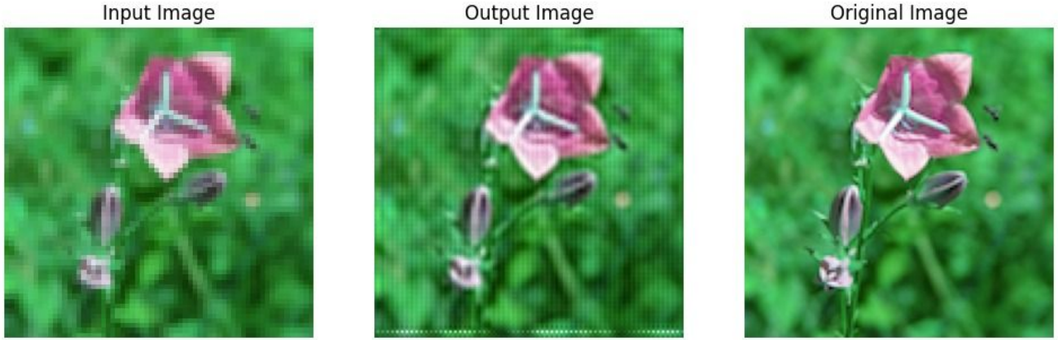
- **Key Takeaways**
  - Haar wavelet showed the highest PSNR and SSIM, indicating better image quality.
  - DB1 and Symlet wavelets exhibited increased noise at the image borders.
  - Based on these results, Haar wavelet was chosen for WaveMixSR, offering improved image reconstruction with minimal artifacts.



Haar Wavelets



DB1 Wavelets



Symlet Wavelets

Results and Discussion (1)

- **Results (PSNR)**
  - VGG11: 31.1 (at 20 epochs)
  - ResNet18: 32.5 (at 20 epochs)

- **Visual Analysis:**
  - ResNet18 produced results that were closer in appearance to the original high-resolution images.
  - VGG11 yielded grainier results compared to ResNet18.

- **Key Takeaways**
  - ResNet18 outperformed VGG11 in both PSNR and visual similarity to the original images.
  - The improved results with ResNet18 suggest its stronger feature extraction capabilities and residual learning benefits.
  - This led to the decision to integrate ResNet18 into the WaveMixSR model, ensuring higher-quality super-resolution.

ResNet18

VGG11

Results and Discussion (2)

- Results
    - **PSNR:**
        - After 20 epochs of adversarial training, PSNR reached 34.46, indicating a significant improvement in image quality.
    - **Discriminator Loss:**
        - True Cross-Entropy Loss: 0.0011899899538150589
        - False Cross-Entropy Loss: 0.00041642620717529 67
- Key Takeaways
    - Adversarial training with the GAN framework significantly improved image quality.
    - The high PSNR and low discriminator loss suggest that the generator achieved better realism and detail reconstruction.

Input Image

Output Image

Original Image

Thank You

Laxmi Sreenivas IMT2020510
Rithvik Ramasani IMT2020543