University Euromed of Fez

EIDIA

Image Restoration using Deep Learning
****

Elaborated by:
Kaoutar Lakdim
Chaimae Biyaye
Anas Belmessid
Monsef Berjaoui
Manal El kassim
Olaya Latoubi

Supervised by:
Dr.Alae Ammour

school year : 2022/2023

May 28, 2024

Acknowledgments

We would like to express our sincere gratitude to our supervisor, Dr. Alae Ammour, for his invaluable support throughout our journey. His patience, expertise, and availability have been of paramount importance in successfully accomplishing this project. His insightful remarks and wise advice have greatly contributed to our learning and professional development. His confidence in our abilities has motivated us to strive for excellence and surpass our limitations.

We would also like to acknowledge his dedication and commitment, which went beyond the role of a supervisor. His passion for the subject matter and willingness to share his knowledge have inspired us and encouraged us to persevere, even in the face of daunting challenges.

Furthermore, we are immensely grateful to Dr. Alae Ammour for his unwavering assistance in tackling various technical issues and enhancing our knowledge in the field of AI, particularly in computer vision. His expertise and guidance have been instrumental in expanding our understanding and skills in this domain.

His profound understanding of the intricacies of computer vision and his ability to convey complex concepts in a clear and concise manner have been invaluable. He has provided us with practical insights and innovative approaches that have significantly improved the quality of our work.

We are truly fortunate to have had the opportunity to work under his supervision and guidance. His mentorship has not only enriched our project but has also empowered us to become more proficient in our chosen field.

Once again, we extend our heartfelt appreciation to Dr. Alae Ammour for his exceptional support, unwavering commitment, and invaluable contributions throughout this project. We are truly grateful for the knowledge and skills we have gained under his tutelage.

Abstract

This report introduces a robust image denoising model called DnCNN (Deep Convolutional Neural Network). DnCNN has achieved significant success in various computer vision tasks, particularly in image denoising. The proposed denoising model leverages a deep network

architecture comprising multiple convolutional layers, batch normalization, and rectified linear unit (ReLU) activation. This architecture enables the model to effectively learn intricate noise patterns and efficiently remove noise from input images. During training, a comprehensive

dataset of noisy-clean image pairs is utilized. The model's parameters are optimized using mean squared error loss as the objective function, and stochastic gradient descent is employed to iteratively update the model and enhance its denoising performance.

Experimental results demonstrate the superior performance of the proposed DnCNN model in reducing image noise. It outperforms existing denoising techniques in terms of peak signal-to-noise ratio (PSNR) and visual quality. Additionally, it successfully preserves essential image details while effectively suppressing noise artifacts.

In conclusion, the DnCNN-based denoising model presented in this report offers a robust and efficient solution for image denoising tasks. Its deep convolutional layers enable accurate noise removal, making it well-suited for various applications, such as medical imaging, surveillance, and photography, where precise image analysis and interpretation depend on high-quality denoising.

Contents

Chapter 1

An In-Depth Exploration of Image Restoration

1. Introduction

1.1 Definition of Image Restoration

Image restoration is the process of improving the quality and visual fidelity of degraded images that have been affected by various factors such as sensor-related degradation, noise, blurring, or compression artifacts. It involves the application of advanced algorithms and mathematical models to recover the true underlying information that may have been corrupted or lost during the image acquisition or transmission process.

The field of image restoration draws upon principles from image processing, computer vision, and signal processing to develop effective techniques for mitigating the effects of degradation and restoring images to their desired state. By analyzing the characteristics of the degradation and employing sophisticated restoration algorithms, image restoration aims to enhance image clarity, reduce noise, remove artifacts, and improve overall visual quality while preserving the important details and features of the original image. This interdisciplinary field continues to evolve with ongoing research and technological advancements, offering valuable solutions for a wide range of applications in fields such as medical imaging, surveillance, remote sensing, and digital photography.

1.2 Overview of the Importance of Image Restoration in Various Domains

Image restoration is very important in many different areas because it can improve and bring back life to damaged or degraded images. One important area where it is used is in preserving historical and cultural items. By restoring faded images, old paintings, and delicate documents, we can make sure that future generations can still see and learn from them. This helps us understand our past and cultural heritage better.

Image restoration is also crucial in forensic investigations. It helps investigators extract important information from blurry or low-quality surveillance footage and crime scene photos. By improving these pictures, they can find valuable evidence and solve difficult cases.

It is also equally important in the fields of art and media. It breathes new life into iconic creations such as vintage movies and photographs by erasing flaws, enhancing colors, and improving overall visual quality. This process not only revitalizes these works but also ensures that their artistic essence is preserved, allowing people to appreciate them now and in the future.

In addition, image restoration plays a significant role in the realm of medical imaging. Diagnostic images, like those obtained from X-rays or MRI scans, may suffer from issues such as noise, artifacts, or poor contrast. By employing restoration techniques, healthcare professionals can enhance the quality of these images, bringing out crucial details and increasing the accuracy of diagnoses. This, in turn, leads to better patient care as medical experts can make well-informed decisions based on precise and reliable visual information.

Figure 1.1: Restore Old Photos Back to Life

1.3 An Exploration of Degradation Sources

Restoration techniques aim to mitigate the effects of degradations or noise encountered in an image. Therefore, it is essential to examine the potential sources of degradation to better understand the challenges involved.

Identifying the sources of degradation is a critical step in image restoration. Various factors can contribute to image degradation, such as sensor noise, motion blur, compression artifacts, and atmospheric interference. By comprehensively exploring these sources, we gain insights into the specific types of degradation present in an image and can tailor our restoration approaches accordingly.

1.3.1 Degradation Related to the Acquisition Context

In this first category, we encounter unexpected events that modify the conditions of signal acquisition. The simplest example is motion blur. We can also envision a sudden change in lighting conditions, resulting in under or overexposure of the observed object. The unpredictable nature of these events makes it nearly impossible to correct their effects without the intervention of a human operator to activate the appropriate restoration process (such as blur correction, motion compensation, or light drift correction).

1.3.2 Degradation Related to the Sensor

Degradation related to the sensor is a crucial aspect affecting the quality of digital images. Sensors are at the forefront of image acquisition, converting light into electrical signals for further processing. However, various factors can introduce degradation during this conversion process. One common challenge is the presence of sensor noise, which can manifest as random variations or patterns in the captured image. Additionally, sensor limitations may lead to reduced dynamic range, resulting in the loss of details in both highlight and shadow regions. Furthermore, sensor artifacts such as color fringing or pixel non-uniformity can distort the true representation of the scene. Understanding and mitigating these sensor-related degradations are essential for achieving high-quality image restoration and ensuring accurate visual perception.

1.4 Types of Image Noise

The following are common types of image noise:

- **Gaussian noise**: Caused by random variations in pixel values, resulting in a subtle grainy pattern.
- **Salt-and-pepper noise**: Randomly occurring black and white pixels scattered throughout the image, resembling grains of salt and pepper.
- **Poisson noise**: Typically encountered in low-light conditions, introducing a granular pattern with varying intensities.
- **Speckle noise**: Appears as granular interference in the image, often resembling a grainy texture.
- **Quantization noise**: Arises from the limited number of levels used to represent pixel values, resulting in a loss of fine details.

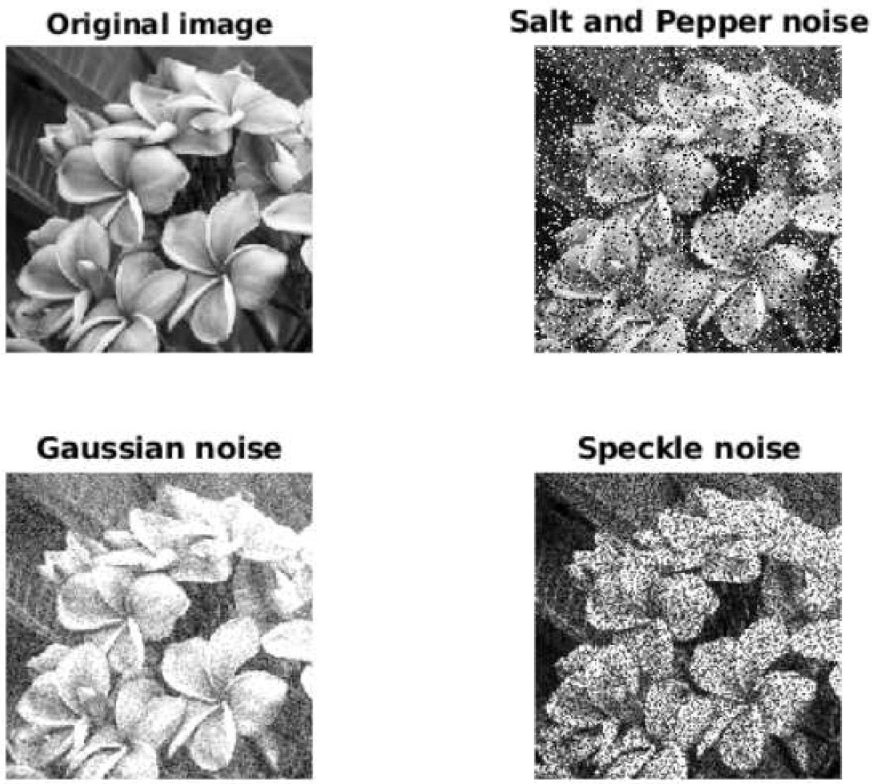


Figure 1.2: Noise Addition

2. Literature Review

2.1 Classical Image Restoration Techniques

- **Noise Reduction**: These techniques aim to reduce unwanted noise in an image.
  - **Median Filtering**: Replaces each pixel with the median value from its neighborhood, which helps remove random noise.
  - **Gaussian Filtering**: Applies a Gaussian blur to smooth out noise while preserving important image details.
  - **Wiener Filtering**: Uses statistical properties of noise and the underlying signal to minimize noise and enhance the image.
- **Deblurring**: These methods address the problem of blurry images caused by various factors.
  - **Inverse Filtering**: Tries to estimate the original image by applying an inverse filter to the blurred image.
  - **Wiener Deconvolution**: Utilizes a Wiener filter to estimate the original image based on knowledge of the blurring process and noise characteristics.
  - **Blind Deconvolution**: Attempts to estimate both the original image and the blurring kernel without prior knowledge.
- **Super-Resolution**: These techniques aim to enhance the resolution and level of detail in an image.
  - **Interpolation-based methods**: Use mathematical interpolation techniques to fill in missing pixels and increase image resolution.
  - **Example-based methods**: Employ a database of high-resolution patches to reconstruct missing details in a low-resolution image.
  - **Regularization-based methods**: Apply regularization techniques to restore high-frequency information and enhance image details.
- **Artifact Removal**: These methods target specific artifacts that occur during image acquisition or compression.
  - **JPEG Artifact Removal**: Focuses on reducing blocky artifacts introduced by JPEG compression.
  - **Salt-and-Pepper Noise Removal**: Aims to remove randomly occurring black and white pixels in an image.
  - **Moiré Pattern Removal**: Addresses interference patterns that appear when photographing or scanning screens or overlapping regular patterns.
- **Image Inpainting**: These techniques fill in missing or damaged regions in an image.
  - **Exemplar-based methods**: Search for similar patches in the surrounding area to fill in missing regions.
  - **PDE-based methods**: Use partial differential equations to propagate information from known regions into the missing areas.
  - **Texture synthesis methods**: Reconstruct missing regions by synthesizing textures based on the image context.

NOM	Date	Dataset	algorithmes	Results
Residual Dense Network for Image Restoration	9, SEPTEMBER 2014	•Set5: This dataset consists of five high-quality images used for image super-resolution. •Set14: This dataset consists of 14 images used for image denoising and super-resolution. •BSD100: This dataset consists of 100 natural images used for image denoising and super-resolution. •Urban100: This dataset consists of 100 high-resolution images of urban scenes used for image super-resolution. •DIV2K: This dataset consists of 900 high-resolution images used for image super-resolution.	• Residual Dense Network (RDN) • Convolutional Neural Networks (CNNs) to improve the performance • data augmentation, • batch normalization, • Adam optimizer comparison 1. Peak Signal-to-Noise Ratio (PSNR), 2. Structural Similarity Index (SSIM), 3. Mean Absolute Error (MAE)	1. Image super-resolution, which are state-of-the-art results. 2. Set14: RDN achieves a PSNR of 29.77 dB and an SSIM of 0.8504 for image denoising, which are state-of-the-art results. 3. BSD100: RDN achieves a PSNR of 28.47 dB and an SSIM of 0.7673 for image denoising, which are state-of-the-art results. 4. Urban100: RDN achieves a PSNR of 26.47 dB and an SSIM of 0.7954 for image super-resolution, which are state-of-the-art results. 5. DIV2K: RDN achieves a PSNR of 30.81 dB and an SSIM of 0.8773 for image super-resolution, which are state-of-the-art results.

Table 1.1: Example Table

NOM	**Date**	**Dataset**	**algorithmes**	**Results**
Deep Learning for Image Super-resolution: A Survey	February 15, 2019	• LSUN: A large-scale dataset of images from different indoor and outdoor scenes, such as bedrooms, kitchens, and churches. • FFHQ: A dataset of human faces that was specifically designed for evaluating generative models. The dataset includes high-quality images with a wide range of poses, expressions, and backgrounds. • Anime: A dataset of anime-style images, which includes different genres and art styles.	1. Mapping Network 2. Noise Injection 3. Progressive Growing 4. Path Length Regularization 5. Adaptive Instance Normalization (AdaIN):	• Improved Image Quality: • Improved Diversity: • Fine-Grained Control: • Interpolation

NOM	Date	Dataset	algorithmes	Results
Learning Image Restoration without Clean Data	March 13, 2018	• Photographic Noise Dataset: The authors used a dataset of photographic images corrupted by various types of noise, including Gaussian, Poisson, and impulse noise. • Monte Carlo Denoising Dataset: The authors used a synthetic dataset of Monte Carlo rendered images corrupted by Poisson noise. • MRI Reconstruction Dataset: The authors used a dataset of undersampled MRI scans, which were reconstructed using a compressed sensing approach that resulted in highly noisy images.	• Neural Network Architecture • Loss Function • Training Data • Training Method • Regularization	•The Noise2Noise method achieves state-of-the-art performance in image restoration using only noisy data for training. •It effectively restores images corrupted by different types of noise (e.g., Gaussian, Poisson, impulse noise). •Compared to methods relying on clean data for training on the Monte Carlo denoising dataset, Noise2Noise outperforms them. •It achieves comparable performance to state-of-the-art methods trained on clean data for photographic noise removal. •Noise2Noise successfully reconstructs high-fidelity, undersampled MRI scans even in the presence of significant noise. •Restoring images from noisy data alone improves performance and simplifies data acquisition.

NOM	Date	Dataset	algorithmes	Results
Deep Image Prior	1, March 2018	The authors use various types of images in their experiments, including natural images (such as landscapes, portraits, and animals) and synthetic images (such as digitized handwritten digits). The images are corrupted by different types of noise and artifacts, such as Gaussian noise, salt-and-pepper noise, missing pixels, and compressed images. The authors also conduct experiments on Magnetic Resonance (MR) images of the human brain, using DIP to denoise the images and compare the results to state-of-the-art methods. The specific datasets used in the experiments are not explicitly stated in the paper.	•Initialize a deep neural network with random weights. • Generate an initial guess for the output image by passing random noise through the neural network. • Compute the loss function between the output image and the target image, which measures the difference between the two images. • Update the neural network parameters using backpropagation to minimize the loss function. • Repeat steps 3-4 for a fixed number of iterations, or until the loss function converge)	• produce competitive or superior results compared to state-of-the-art methods • demonstrate that the DIP method can handle a wide range of different corruptions, such as Gaussian noise, salt-and-pepper noise, • the DIP method can be used for medical image denoising, specifically Magnetic Resonance (MR) images of the human brain, and achieve competitive results compared to state-of-the-art methods. • beyond image restoration, such as tomography and microscopy, and could be particularly useful in scenarios where training data is scarce or unavailable

NOM	Date	Dataset	algorithmes	Results
An Old Photo Image Restoration Processing Based on Deep Neural Network Structure	21 January 2023	The dataset used in this research article for restoring old photos is not explicitly mentioned. However, it is mentioned that the dataset includes both degraded and high-quality versions of the same images.		The results show that the peak signal-to-noise ratio and structural similarity of the algorithm in this paper are higher than those of other algorithms. Therefore, compared with other algorithms, the algorithm in this paper not only has better performance in blur repair but also has better performance in damage repair. The model in this article is more suitable for the field of restoration of old photos. From the experimental point of view, the restoration effect is good, but there is still improvement in the restoration effect, especially the restoration of damaged photos which is not ideal enough, so the focus of future work will be put on repairing broken photos.

NOM	Date	Dataset	algorithmes	Results
On-Demand Learning for Deep Image Restoration	October 2017.Venice, Italy.	they used the Berkeley Segmentation Dataset (BSDS) to train the deep learning models that were used as the restoration model in their experiments.	stochastic gradient descent (SGD)	the algorithm achieved a peak signal-to-noise ratio (PSNR) of 33.83 dB on the BSDS500 test set
Image Restoration Using Deep Learning	2016,Belgium	•The BSDS500 dataset: This dataset contains 500 natural images that have been manually annotated with ground-truth segmentation maps. It is commonly used for evaluating algorithms for tasks such as denoising and super-resolution. •The Set5 and Set14 datasets: These datasets contain 5 and 14 high-resolution images, respectively, that are commonly used for evaluating super-resolution algorithms. •The GoPro dataset: This dataset contains a large number of blurry and sharp image pairs, which are used for evaluating deblurring algorithms.	the deep denoising prior (DDP) network	•Denoising: The DDP network achieves a peak signal-to-noise ratio (PSNR) of 33.9 dB on the BSDS500 dataset, outperforming several other state-of-the-art denoising algorithms. •Super-resolution: The DDP network achieves a PSNR of 32.62 dB on the Set5 dataset and 28.40 dB on the Set14 dataset, outperforming several other state-of-the-art super-resolution algorithms. •Deblurring: The DDP network achieves a PSNR of 29.20 dB on the GoPro dataset, outperforming several other state-of-the-art deblurring algorithms.

NOM	Date	Dataset	algorithmes	Results
Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections	was first published on arXiv on June 28, 2016, and the most recent version (version 3) was published on January 11, 2017.	The Berkeley Segmentation Dataset (BSDS300) The Set12 and Set68 datasets The Urban100 dataset The DIV2K dataset	•Filtering techniques: Inpainting technique •Super-resolution techniquesues •Deconvolution techniques: Deep learning techniques	The article proposes a novel approach to image restoration using symmetric skip-connected convolutional autoencoders. It outperforms networks without skip connections or with asymmetric skip connections. The approach is robust to various degradation levels and can handle denoising, deblurring, and inpainting tasks. It surpasses state-of-the-art methods like DRN and FCN, as confirmed by extensive performance analysis on different datasets and degradation types. Overall, the proposed approach with symmetric skip connections achieves high-quality image restoration across diverse degradation types.

NOM	Date	Dataset	algorithmes	Results
Literature Review on Image Restoration	December 2022	•CelebA •SVHN •Cars	•CVAE (Conditional Variational Autoencoder) •CGAN (Conditional Generative Adversarial Network) •FM-CGAN (Feature-Matching Conditional Generative Adversarial Network) •CVAE-GA (Conditional Variational Autoencoder with Generative Adversarial Networks)	•CVAE : 8.09 •CGAN :61.97 •FM-CGAN : 79.76 •CVAE-GA : 97.78 •-CVAE-GAN performed better than other algorithms, and its accuracy was very close to real data.
Image Restoration Using a Neural Network	7, JULY 1988	"Image Restoration Using a Neural Network" tests the proposed restoration algorithm on synthetic and real images. Synthetic images are generated using a known degradation model, while real images undergo convolution with a blur function and the addition of white Gaussian noise. The paper also mentions the use of Laplacian thresholding to select sample pixels from the images. No specific datasets are mentioned.	1. A dynamic iterative algorithm is used to generate images by updating gray level neurons through a simple sum scheme. It effectively reconstructs the image after estimating the parameters of the neural network model. 2. Efficient image restoration algorithm reduces complexity, matches original results. Faster, fault-tolerant, uses multilevel approach and secondary neural network.	

2.2 Advances in deep learning for image restoration

1. **Attention Mechanisms:** Attention mechanisms have been introduced to enhance the performance of deep learning models for image restoration. These mechanisms allow the network to focus on relevant image regions while suppressing noise or artifacts. Self-attention modules, such as those used in the Transformer architecture, have been integrated into image restoration models to improve their ability to capture long-range dependencies and contextual information.

2. **Self-supervised Learning:** Deep learning models for image restoration often require large amounts of labeled data, which can be expensive and time-consuming to obtain. Self-supervised learning techniques alleviate this limitation by utilizing the inherent structure or content within the image data itself for training. For example, image denoising can be treated as a self-supervised task where a network learns to predict clean patches from noisy patches.

3. **Multi-modal and Multi-scale Approaches:** To handle complex image restoration tasks, deep learning models have incorporated multi-modal and multi-scale approaches. Multi-modal methods exploit multiple sources of information, such as combining infrared and visible light images, to enhance restoration performance. Similarly, multi-scale models process images at different resolutions or scales, allowing them to capture both fine details and global structures effectively.

4. **Unsupervised Learning:** While labeled datasets are valuable, unsupervised learning methods aim to restore images without the need for paired clean and degraded examples. By utilizing unsupervised learning, models can learn from unpaired data, such as uncorrupted images and their corresponding degraded versions, resulting in more flexible and adaptable image restoration algorithms.

5. **Domain Adaptation and Transfer Learning:** Deep learning models trained on large-scale datasets may not perform well when applied to specific domains or real-world scenarios. Domain adaptation and transfer learning techniques address this issue by fine-tuning pre-trained models on smaller or more specific datasets, allowing them to generalize better to target domains and improve restoration performance.

3. State-of-the-Art in Deep Learning-Based Image Restoration

. Recent Techniques and Trends:

Deep learning based image restoration has emerged as the forefront of computer vision, exhibiting groundbreaking advancements in enhancing image quality. With the rapid evolution of deep neural networks and increased computational power, **state-of-the-art techniques** have achieved unparalleled results across a range of image restoration tasks.

Among the notable applications of deep learning in this domain is **image denoising**. Traditional denoising methods often struggle to effectively remove noise while preserving intricate details. However, deep learning approaches have demonstrated exceptional performance by leveraging large-scale datasets and convolutional neural networks (CNNs). By exploiting the underlying structure and patterns of clean images, these models effectively suppress noise while maintaining image fidelity, resulting in visually appealing denoised outputs.

Another critical area where deep learning has made significant strides is **image super-resolution**. This technique aims to enhance the resolution and sharpness of low-resolution images. Deep learning-based methods employ sophisticated architectures, such as residual blocks and skip connections, within CNNs to learn the intricate mapping between low-resolution and high-resolution image pairs. By capturing the inherent correlation between these pairs, these models generate visually compelling and highly detailed high-resolution images.

Moreover, deep learning has also found success in **image inpainting**, which involves filling in missing or damaged regions within an image. These techniques capitalize on contextual information from surrounding regions to accurately infer the missing content. By training deep neural networks on extensive datasets of complete images, these models effectively learn to predict missing regions with realistic and visually coherent outcomes.

In addition to denoising, super-resolution, and inpainting, deep learning-based approaches have extended their prowess to other image restoration tasks, including image deblurring, image dehazing, and image colorization. These methods harness advanced network architectures, sophisticated loss functions, and refined training strategies to achieve cutting-edge outcomes.

Chapter 2

Theoretical Basis

1. Multilayer Perceptron

Input Layer: The input layer of the MLP receives the preprocessed image as input. Each neuron in the input layer represents a pixel or a group of pixels in the image.

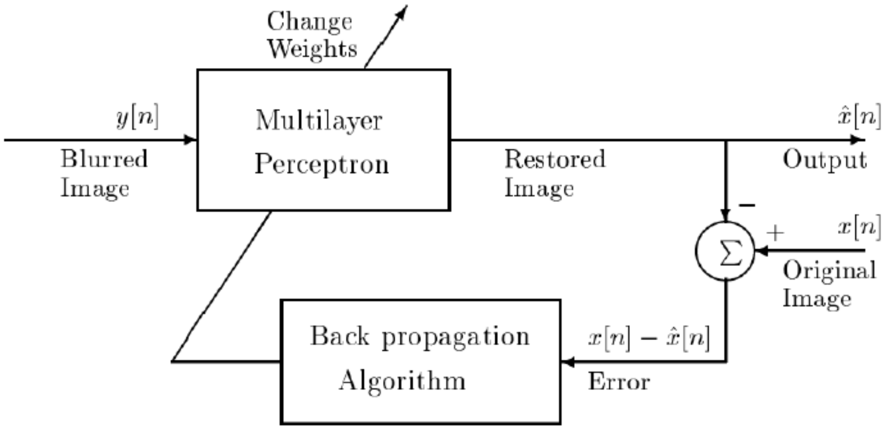


Figure 3: Multilayer Perceptron Model for Image Restoration

Architecture:

1. **Hidden Layers:** The hidden layers of the MLP perform computations on the input data to learn relevant features and patterns. Each neuron in the hidden layers applies a linear transformation to the inputs followed by a non-linear activation function (e.g., sigmoid or ReLU). These transformations help the MLP model learn complex relationships and extract meaningful information from the image.

2. **Loss Functions:** Loss functions quantify the discrepancy between the predicted restored image and the ground truth image. They provide a measure of how well the MLP is performing in terms of image restoration. In image restoration tasks, common loss functions include mean squared error (MSE), which calculates the average squared difference between the predicted and ground truth pixel values, and structural similarity index (SSIM), which measures the similarity between the predicted and ground truth images based on luminance, contrast, and structural information. The choice of the loss function depends on the specific requirements and characteristics of the restoration task

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$$SSIM(x,y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2}\frac{2\sigma_{xy}}{\sigma_x^2 + \sigma_y^2}$$

with:

- μ_x the average of x;
- μ_y the average of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;
- $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator;
- L the dynamic range of the pixel-values (typically this is $2^{\#bits\ per\ pixel} - 1$);
- $k_1 = 0.01$ and $k_2 = 0.03$ by default.

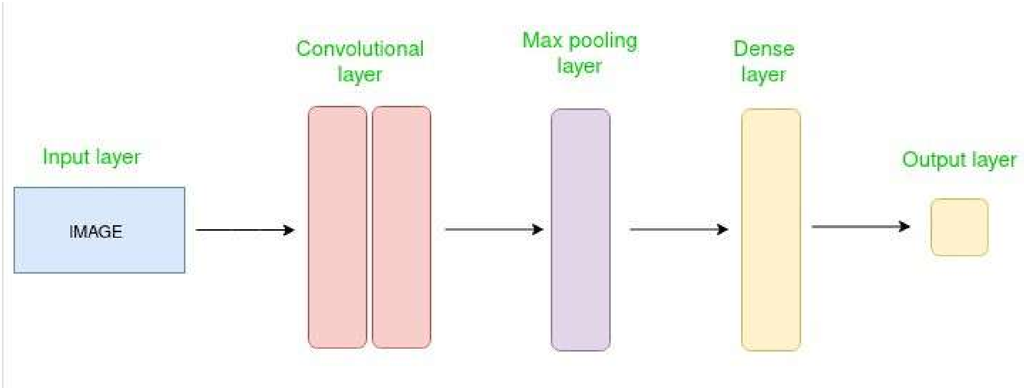
3. **Optimization Algorithms:**Optimization algorithms are used to update the weights of the MLP during the training process, with the goal of minimizing the loss function and improving the model's performance. Gradient descent is a widely used optimization algorithm for training MLPs. It calculates the gradients of the loss function with respect to

the model's parameters (weights and biases) and adjusts these parameters in the direction that reduces the loss. Other variants of gradient descent, such as stochastic gradient descent (SGD) or Adam optimizer, can also be employed to accelerate the training process and handle larger datasets

4. **Output Layer:** Deep learning models for image restoration often require large amounts of labeled data, which can be expensive and time-consuming to obtain. Self-supervised learning techniques alleviate this limitation by utilizing the inherent structure or content within the image data itself for training. For example, image denoising can be treated as a self-supervised task where a network learns to predict clean patches from noisy patches.

2. Convolutional Neural Network

A Convolutional Neural Network (CNN): is a type of Deep Learning neural network architecture commonly used in Computer Vision. Computer vision is a field of Artificial Intelligence that enables a computer to understand and interpret the image or visual data.



The Convolutional layer applies filters to the input image to extract features, the Pooling layer downsamples the image to reduce computation, and the fully connected layer makes the final prediction. The network learns the optimal filters through backpropagation and gradient descent.
Architecture:

1. **Input Data:** The input data for a CNN is typically a collection of images or image-like data. Each image is represented as a multidimensional array of pixel values, with dimensions for width, height, and color channels.

2. **Convolutional Layers:** The input images are passed through one or more convolutional layers. Each convolutional layer applies a set of filters to the input image, producing a set of feature maps. These filters capture different patterns and local structures in the image.

3. **Activation Function:** After each convolutional layer, an activation function is applied element-wise to the feature maps. The most common activation function used in CNNs is the Rectified Linear Unit (ReLU), which introduces non-linearity and helps the network learn complex representations.

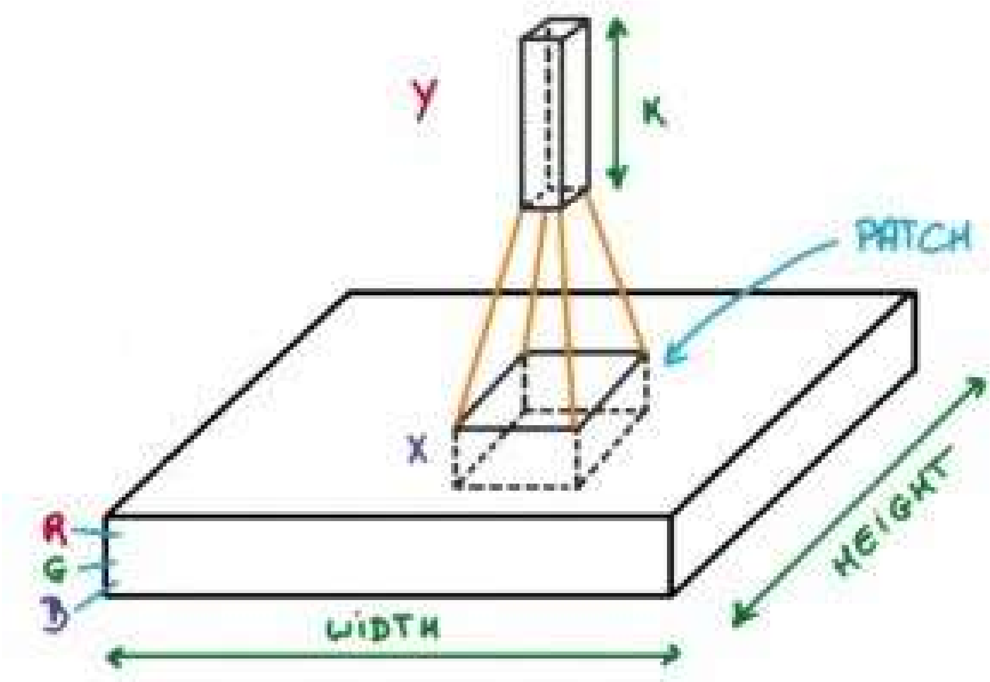
4. **Pooling Layers:** The feature maps obtained from the convolutional layers are often downsampled using pooling layers. Pooling helps reduce the spatial dimensions of the data while retaining important features. The most common type of pooling is max pooling, which selects the maximum value within each pooling region.

5. **Fully Connected Layers:** After several convolutional and pooling layers, the resulting feature maps are flattened into a one-dimensional vector. This vector is then fed into one or more fully connected layers. These layers are responsible for learning high-level representations and making predictions based on the extracted features.

6. **Activation and Dropout:** Activation functions, such as ReLU, are applied to the outputs of the fully connected layers. Additionally, dropout regularization may be used to randomly deactivate a certain percentage of neurons during training, preventing overfitting.

7. **Output Layer:** The final fully connected layer is typically connected to an output layer, which has a number of neurons equal to the number of classes in the classification task. The output layer applies an activation function appropriate for the task, such as softmax for multi-class classification, to generate class probabilities.

How Convolutional Layers works?



Convolution layers consist of a set of learnable filters (or kernels) having small widths and heights and the same depth as that of input volume (3 if the input layer is image input).

For example, if we have to run convolution on an image with dimensions 34x34x3. The possible size of filters can be axax3, where 'a' can be anything like 3, 5, or 7 but smaller as compared to the image dimension.

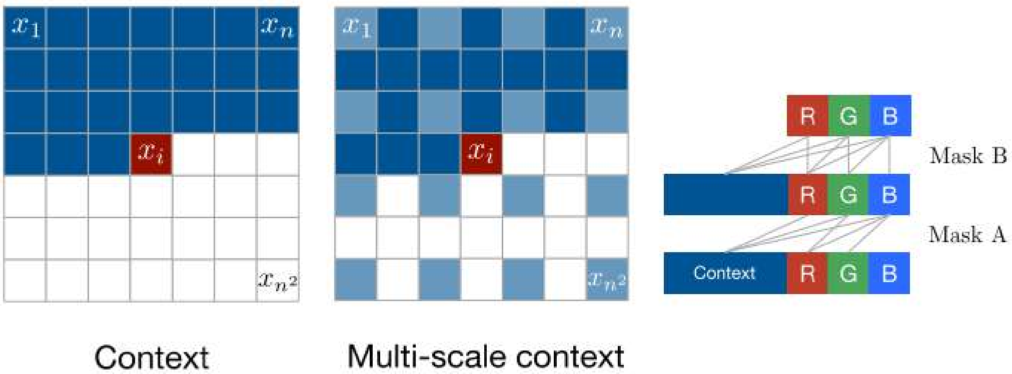
During the forward pass, we slide each filter across the whole input volume step by step where each step is called stride (which can have a value of 2, 3, or even 4 for high-dimensional images) and compute the dot product between the kernel weights and patch from input volume.

As we slide our filters we'll get a 2-D output for each filter and we'll stack them together as a result, we'll get output volume having a depth equal to the number of filters. The network will learn all the filters.

3. PixNet

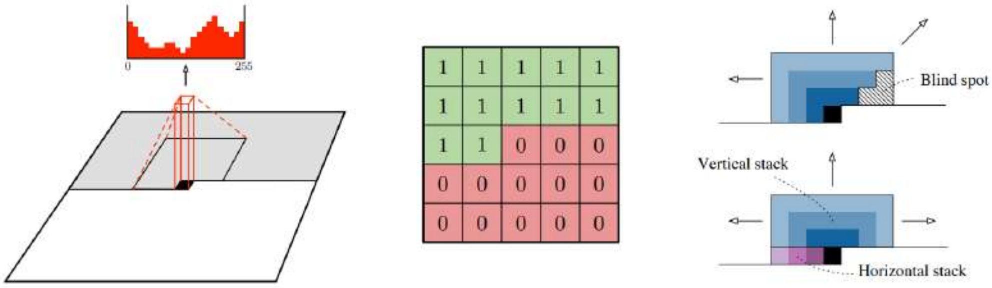
A Pixel CNN-like network: refers to a type of generative model that is inspired by the Pixel CNN architecture. The Pixel CNN (Pixel-Recursive Super Resolution) is a convolutional neural network (CNN) designed for image generation, proposed by van den Oord et al. in 2016.

The Pixel CNN model is called "pixel-recursive" because it generates images pixel by pixel in a recursive manner. It takes the previously generated pixels as input to predict the current pixel, thus capturing the dependencies between neighboring pixels. This autoregressive approach allows the model to generate high-quality images with fine-grained details. .



Architecture:

1. **Input Layer:** The network takes an image or a partial image as input.
2. **Convolutional Layers:** The convolutional layers form the core of the Pixel CNN-like network. They are responsible for capturing the dependencies between pixels. Each convolutional layer applies a set of filters to the input image to extract features.
3. **Masked Convolutions:**In order to maintain the autoregressive property, where each pixel is generated conditioned on previously generated pixels, masked convolutions are used. Masking ensures that each pixel only has access to the pixels already generated or to its left and above, preventing any "future" information leakage. There are different types of masks used, such as a horizontal mask or a fully visible mask, depending on the specific variant of the Pixel CNN-like network.
4. **Non-Linearity:** Activation functions like ReLU (Rectified Linear Unit) or Leaky ReLU are typically used to introduce non-linearity and enhance the expressive power of the network.
5. **Residual Connections:** To help with the flow of gradients during training and ease the optimization process, residual connections can be employed. These connections enable the network to bypass certain layers and directly propagate information from earlier layers to subsequent layers.
6. **Output Layer:** To help with the flow of gradients during training and ease the optimization process, residual connections can be employed. These connections enable the network to bypass certain layers and directly propagate information from earlier layers to subsequent layers.



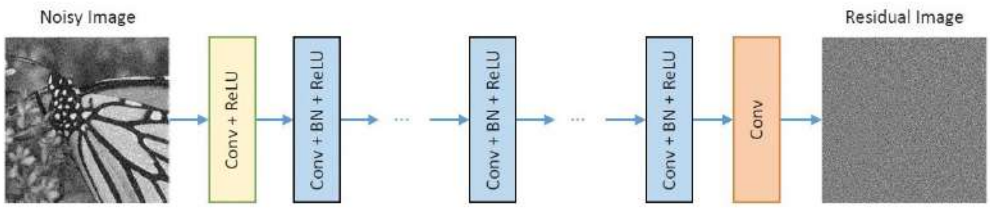
Left: A visualization of the PixelCNN that maps a neighborhood of pixels to prediction for the next pixel. To generate pixel xi the model can only condition on the previously generated pixels .

Middle: an example matrix that is used to mask the 5x5 filters to make sure the model cannot read pixels below (or strictly to the right) of the current pixel to make its predictions.

Right: Top: PixelCNNs have a blind spot in the receptive field that can not be used to make predictions,Bottom: Two convolutional stacks (blue and purple) allow to capture the whole receptive field.

4. DnCNN

DnCNN: stands for "Denoising Convolutional Neural Network." It is a specific type of CNN architecture designed for image denoising, which is the task of removing noise from a corrupted or noisy image to obtain a cleaner version. .



Architecture:

1. **Input:**The input to the DnCNN is a noisy image. The noisy image is typically represented as a grayscale image with pixel values ranging from 0 to 255.

2. **Convolutional Layers:** DnCNN typically consists of multiple convolutional layers. These layers perform convolutions on the input image to extract features and learn representations that can differentiate between noise and signal.

3. **Batch Normalization and ReLU:**After each convolutional layer, batch normalization is often applied. Batch normalization normalizes the outputs of each layer, making the network more robust and aiding in the training process. Additionally, the Rectified Linear Unit (ReLU) activation function is applied to introduce non-linearity.

4. **Residual Learning:** One distinctive aspect of DnCNN is the use of residual learning. Residual learning involves using skip connections that bypass some layers and directly connect the input to the output. These connections help preserve important image details during the denoising process and facilitate the training of deeper networks.

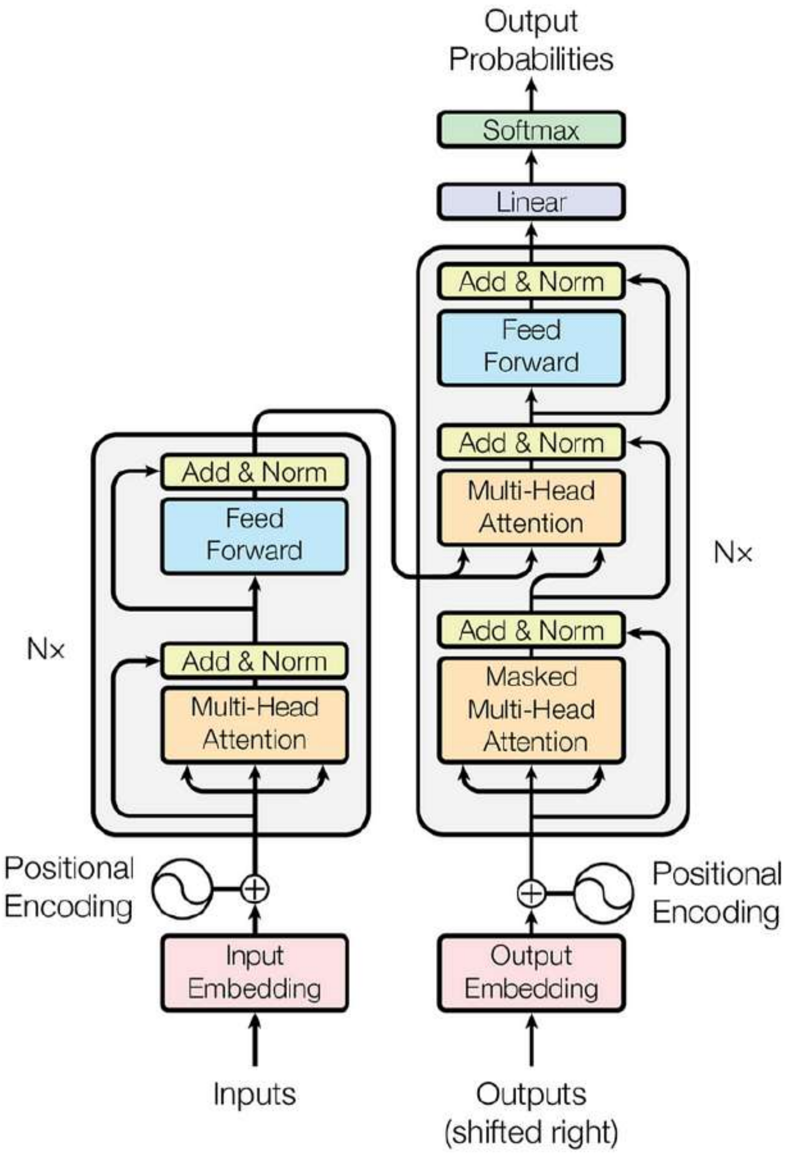
5. **Convolutional Layers (again)**After the skip connections, additional convolutional layers are applied to refine the features and learn more complex representations of the noisy image.

6. **Output Layer:** The final output of the DnCNN is a denoised image, which aims to remove the noise while retaining the important structures and details of the original image.

5.The Transformer

Transformer: Transformers are a type of deep learning model architecture that has gained significant attention and revolutionized various natural language processing (NLP) tasks. They were introduced in a paper titled "Attention Is All You Need" by Vaswani et al. in 2017 The key idea behind transformers is the self-attention mechanism, which allows the model to weigh the importance of different words or tokens in a sequence when processing the input data. This attention mechanism enables transformers to capture the contextual relationships between words efficiently.

.



Architecture:

1. **Self-Attention:** Self-attention, also known as scaled dot-product attention, allows the model to compute attention weights for each word/token in a sequence based on its relationship with other words/tokens in the same sequence. It captures the dependencies between words by assigning higher weights to more relevant words. The attention weights are used to compute a weighted sum of the input representations, creating context-aware representations for each word/token.

2. **Multi-Head Attention:** Multi-Head Attention: Transformers use multi-head attention, where multiple self-attention heads operate in parallel. Each attention head focuses on different parts of the input, allowing the model to capture different types of relationships and information

3. **Encoder-Decoder Architecture:**Encoder-Decoder Architecture: Transformers often employ an encoder-decoder architecture. In this setup, the input sequence is first processed by an encoder, which consists of multiple layers of self-attention and feed-forward neural networks. The encoder learns contextual representations of the input sequence. Then, a decoder is used to generate an output sequence based on the encoder's representations. The decoder also utilizes self-attention layers but additionally incorporates encoder-decoder attention layers to attend to the relevant parts of the input sequence

4. **Positional Encoding:**Since transformers do not have an inherent notion of word order, positional encoding is used to provide positional information to the model. Positional encodings are added to the input embeddings and allow the model to understand the order of words in the sequence.

5. **Feed-Forward Networks**Transformers employ feed-forward neural networks to process the self-attention outputs. These networks consist of multiple fully connected layers and aim to capture more complex interactions between words..

6. **Residual Connections and Layer Normalization:**To facilitate training deep models, transformers employ residual connections, where the input of a layer is combined with the layer's output. Layer normalization is then applied to the combined results, ensuring stable training and improving the flow of gradients.

6.Peak signal-to-noise ratio

The term peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. Because many signals have a very wide dynamic range, (ratio between the largest and smallest possible values of a changeable quantity) the PSNR is usually expressed in terms of the logarithmic decibel scale The mathematical representation of the PSNR is as follows.

.

$$PSNR = 10\log_{10}\frac{255^2}{MSE}\,\mathrm{dB}$$

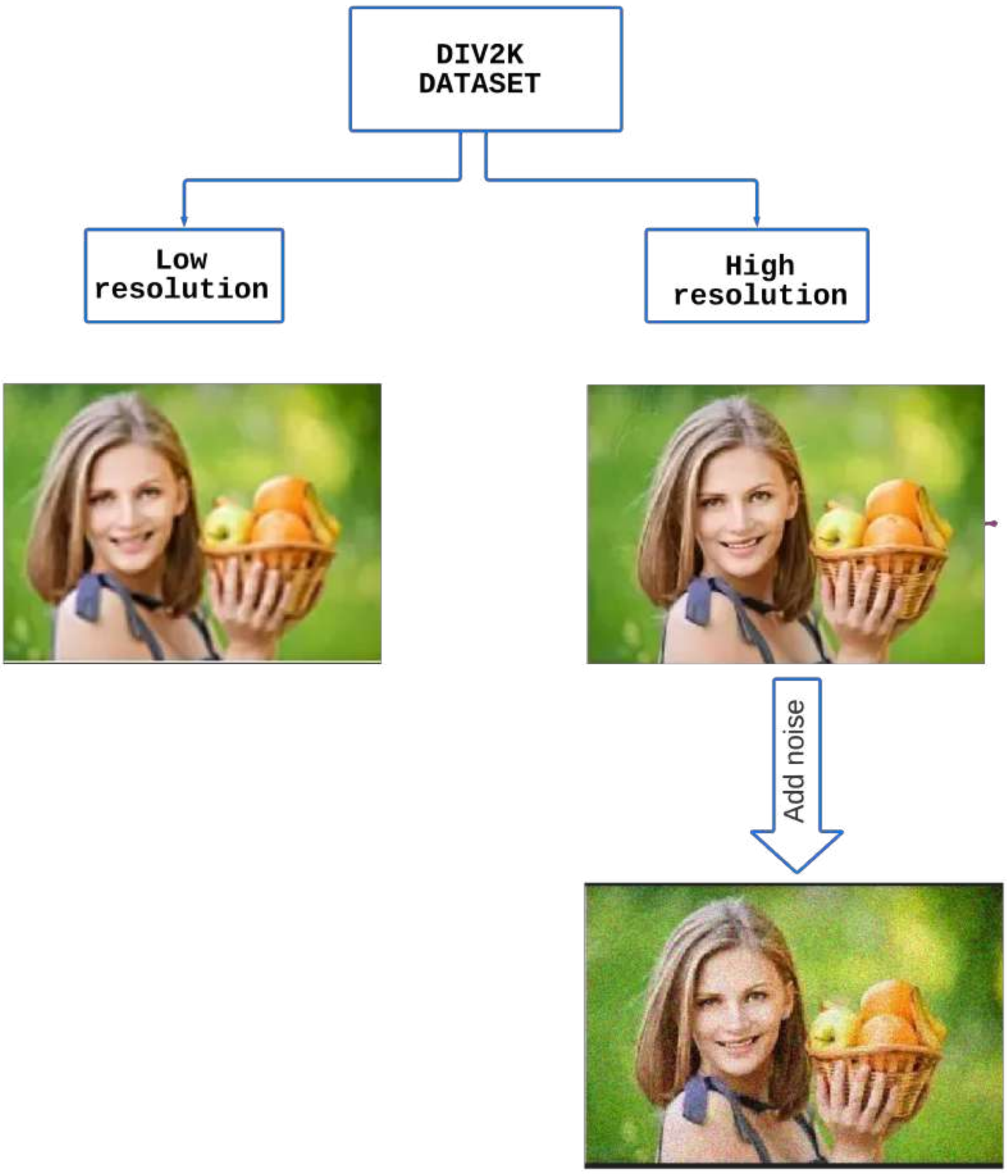
in simpler terms, PSNR computes the ratio between the maximum possible pixel value and the average squared difference between corresponding pixels in the original and reconstructed images. A higher PSNR value indicates a higher similarity between the two images, implying lower distortion or noise.

PSNR is expressed in decibels (dB) and provides a relative measure of image quality. Higher PSNR values typically indicate better quality or less distortion, while lower values indicate more noise or distortion in the reconstructed image compared to the original

Chapter 3

Methodology

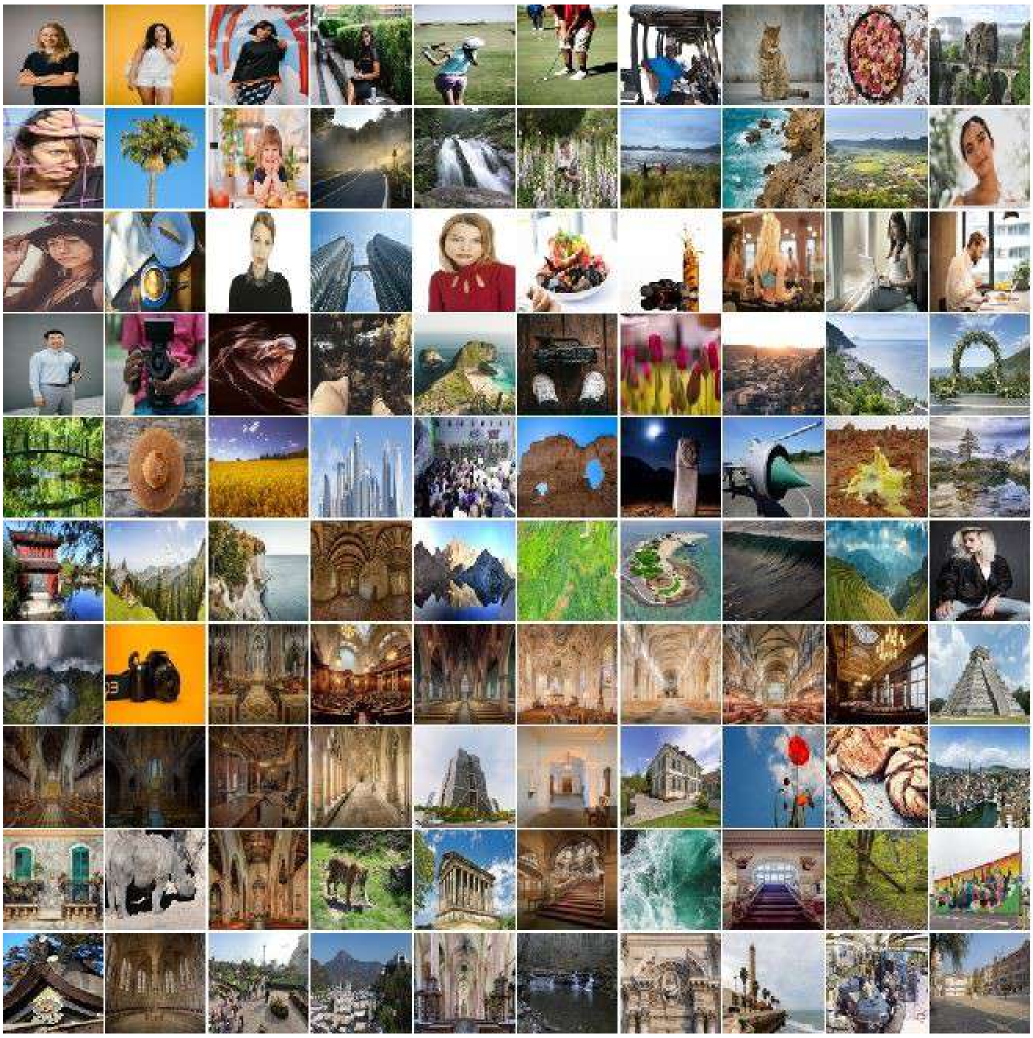
1. Data Description



The DIV2K dataset is a curated collection of high-quality images obtained from various sources. It is a widely used benchmark dataset for image restoration tasks, particularly in the field of single-image super-resolution. Our goal in working with this dataset was to capture various types of scenes, textures, and objects, ensuring coverage of a wide range of challenges commonly encountered in image restoration tasks. This would facilitate research and evaluation of various image restoration algorithms, allowing us to compare and analyze the performance of different approaches.

Origin:

The DIV2K dataset was assembled and made publicly available by a group of researchers and experts in the field of computer vision and image processing. The dataset's creation involved a collaborative effort from multiple contributors with the aim of providing a comprehensive resource for image restoration research.



The images in the DIV2K dataset were collected from various sources, including but not limited to online platforms, databases, and open-access image repositories. The dataset's creators conducted extensive searches and utilized their expertise to curate a diverse collection of high-quality images.

To ensure a broad representation of real-world scenarios, the dataset encompasses different types of scenes, textures, and objects. This careful selection process involved considering a range of factors such as lighting conditions, environmental settings, and object variations. The goal was to capture a wide spectrum of challenges commonly encountered in image restoration tasks.

It is important to note that the images in the DIV2K dataset are a mixture of both professionally photographed images and images obtained from other sources. The professionally photographed images were captured using high-resolution cameras under controlled conditions, ensuring the highest level of visual quality and detail. The other images obtained from various sources went through a rigorous quality assessment to ensure they met the dataset's standards for inclusion.

Size and resolution:

The DIV2K dataset has a total of 1,000 high-quality images used for the NTIRE2017 and NTIRE2018 Super-Resolution Challenges. The dataset is divided into three subsets: 800 images for training, 100 images for validation, and 100 images for testing. These images depict various scenes and encompass different types of degradations to simulate real-world conditions.

The dataset includes low-resolution images with different degradation settings, such as unknown x4 downscaling, realistic mild $\times 4$ conditions with motion blur, Poisson noise, and pixel shifting, and realistic wild $\times 4$ conditions with varying levels of degradation across images.

The images in the dataset possess a level of detail and clarity that goes beyond standard resolution. With resolutions often exceeding 2K (2048x1080) or even 4K (3840x2160), the dataset provides us with an abundance of fine-grained information to work with. This high level of detail allows for the exploration and development of advanced image restoration algorithms that can effectively handle intricate features and preserve visual fidelity.

Partitions: After collecting the DIV2K 1000 images the authors computed image entropy, bit per pixel (bpp) PNG compression rates and CORNIA scores and applied bicubic downscaling $\times 3$ and then upscaling $\times 3$ with bicubic interpolation (imresize Matlab function), ANR [47] and A+ [48] methods and default settings.

Ground Truth Images:

In the DIV2K dataset, each degraded image is accompanied by a corresponding ground truth image. These ground truth images serve as the clean, undistorted references that represent the ideal restoration outcome. They provide researchers with a reliable benchmark to assess the effectiveness and quality of image restoration algorithms.

The ground truth images in the DIV2K dataset are meticulously generated to accurately depict the original, pristine version of the degraded image. They serve as a reference point for comparison, allowing researchers to objectively evaluate the performance of various image restoration techniques.

By having access to the ground truth images, we can quantitatively measure the success of their restoration algorithms. We can employ objective evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to compare the restored images against their corresponding ground truth versions. These metrics provide numerical measures of the similarity between the restored image and the ground truth image in terms of visual quality, detail preservation, and overall fidelity.

Overall, the provision of ground truth images in the DIV2K dataset enhances the reliability, objectivity, and reproducibility of image restoration research. It enables researchers to evaluate and compare the performance of different algorithms, fosters innovation in the field, and ultimately contributes to the development of more advanced and accurate image restoration techniques.

Degradations :

We implemented a noise addition process on the ground truth images in the DIV2K dataset to create more realistic and challenging conditions for evaluating image restoration algorithms. Our team utilized an error diffusion dithering method to introduce noise into the clean ground truth images.

In the noise addition process, we first converted the ground truth images to the RGB color space and loaded the pixel values. We then iterated through each pixel in the image to apply the noise. For each pixel, we used an apply threshold function to introduce noise to the red, green, and blue color channels.

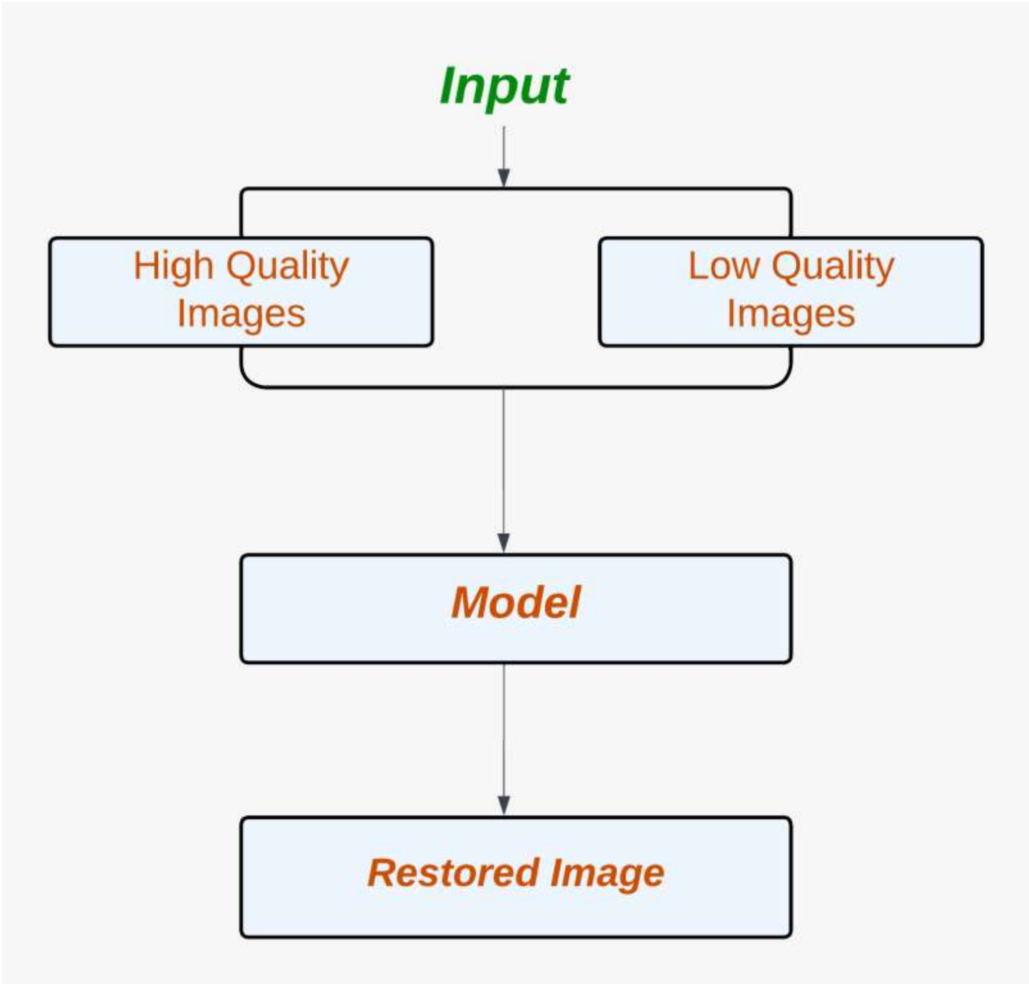
To further enhance the realism of the noise, we employed an error diffusion dithering technique. This method involved propagating the quantization error from the modified pixel to neighboring pixels. Specifically, we calculated the error by subtracting the newly introduced pixel value from the original pixel value. We then distributed this error to adjacent pixels according to specific error diffusion coefficients.

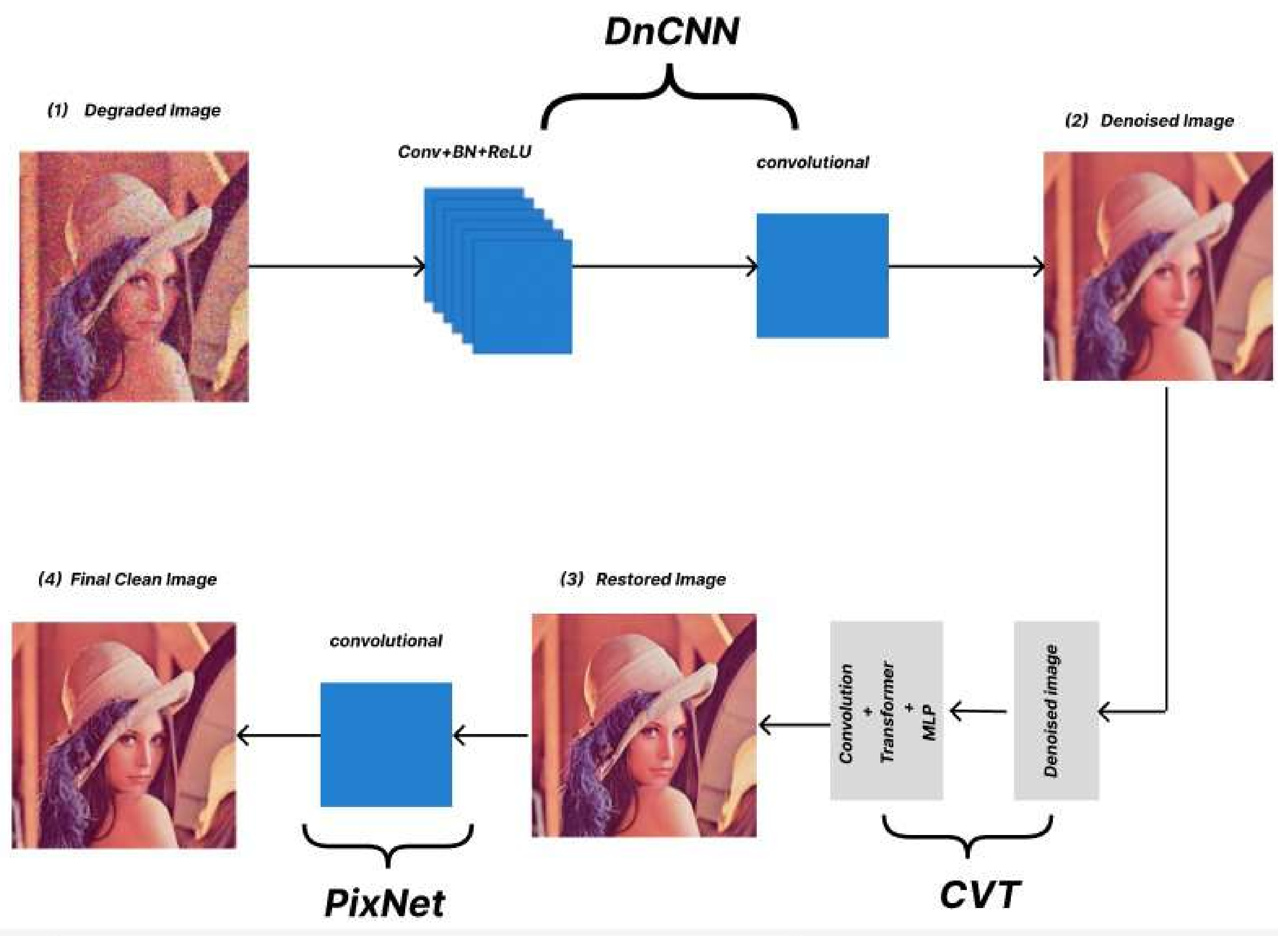
By applying the error diffusion dithering method, we ensured that the noise was spread across neighboring pixels, resulting in a more realistic and visually coherent representation of the added noise. This approach allowed us to simulate the characteristics of various types of noise, such as random patterns, texture variations, and pixel-level irregularities.

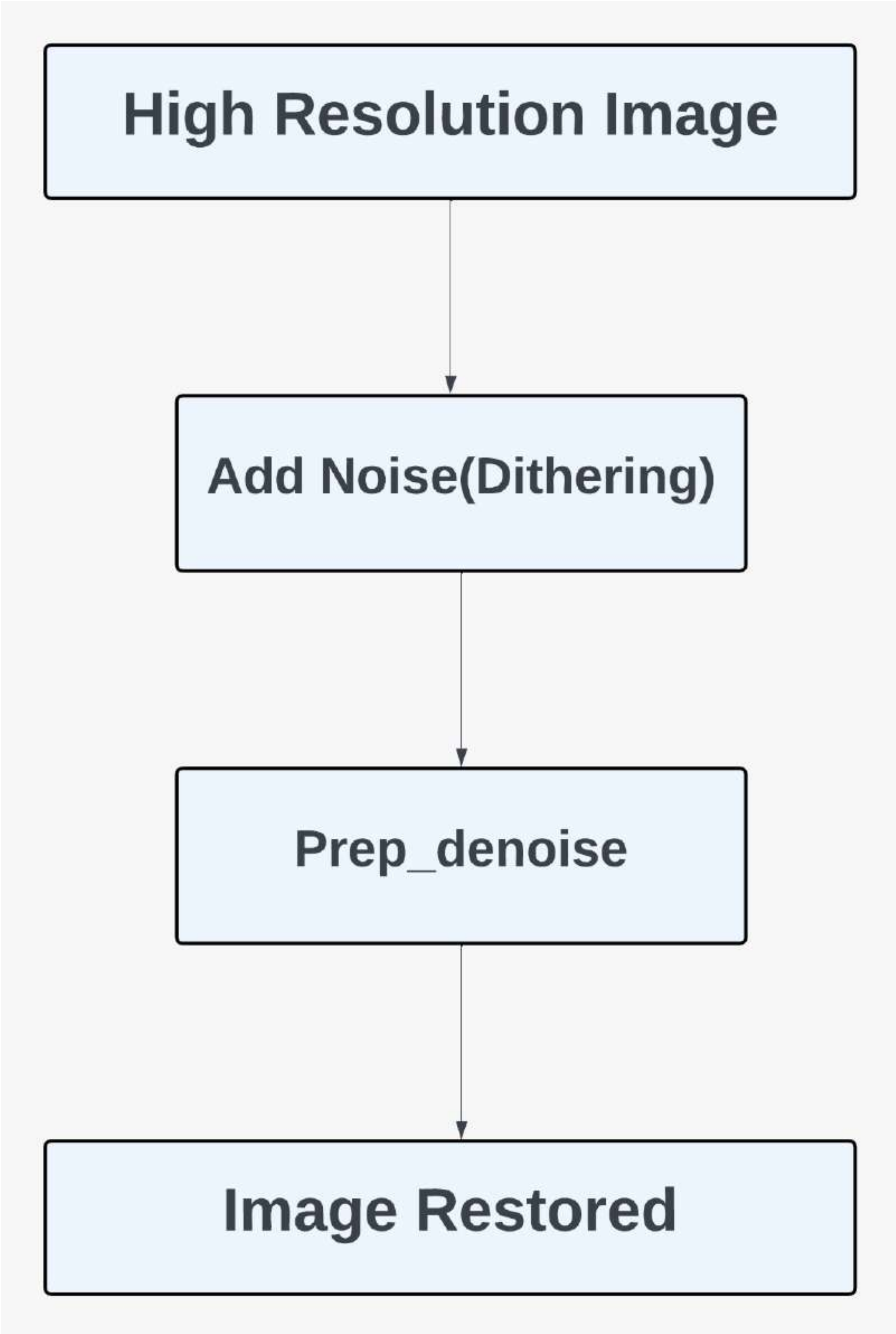
The utilization of the error diffusion dithering method in the noise addition process aimed to create a more challenging and representative dataset for evaluating image restoration algorithms. Researchers can now explore and develop techniques that effectively address the specific types of noise introduced using this method.

2. Model Architecture

After completing the crucial steps of data collection and preparation, the next essential phase in developing an image restoration model is to determine the optimal model architecture. The model architecture serves as the blueprint that outlines how the neural network will process the input images and extract features to restore or enhance them. The choice of model architecture plays a critical role in the model's ability to effectively restore image quality, handle noise, and address various restoration challenges


Input
High Quality Images
Low Quality Images
Model
Restored Image


DnCNN
(1) Degraded Image
Conv+BN+ReLU
convolutional
(2) Denoised Image
(4) Final Clean Image
(3) Restored Image
convolutional
Convolution + Transformer + MLP
Denoised image
PixNet
CVT



3. Model

In the methodology chapter, we delve into the various components and techniques that are meticulously considered and integrated to construct a cohesive architecture for our image restoration model. These components play crucial roles in enhancing the model's ability to restore image quality, handle noise, and address various restoration challenges. Let's explore each component in detail:

3.1 DnCNN (Denoising Convolutional Neural Network)

DnCNN, which stands for Denoising Convolutional Neural Network, is a specialized type of CNN that is specifically design ed for denoising tasks. It is widely used for reducing noise artifacts and enhancing image quality in image restoration applications.

The architecture of DnCNN is characterized by its deep structure, consisting of multiple layers of convolutional operations. These convolutional layers are responsible for learning filters that can effectively suppress noise while preserving important image details. By employing a deep architecture, DnCNN is capable of capturing complex and abstract representations of noise patterns, enabling it to achieve high denoising performance.

One distinctive feature of DnCNN is the use of skip connections. Skip connections allow information from earlier layers to bypass subsequent layers and be directly fed into deeper layers. This enables DnCNN to effectively propagate both low-level and high-level features throughout the network, promoting information flow and helping preserve important image details during the denoising process.

During training, DnCNN is trained on pairs of noisy and clean images. The model learns to minimize the difference between the denoised output and the corresponding clean image, effectively learning to remove noise while preserving the underlying structure and details of the image.

In the context of image restoration, DnCNN can be utilized as a powerful tool to reduce noise artifacts and enhance image quality. By applying DnCNN to noisy images, the model can effectively remove various types of noise, such as Gaussian noise, salt-and-pepper noise, or random noise patterns. The denoising capabilities of DnCNN contribute to improving the overall visual quality and fidelity of restored images.

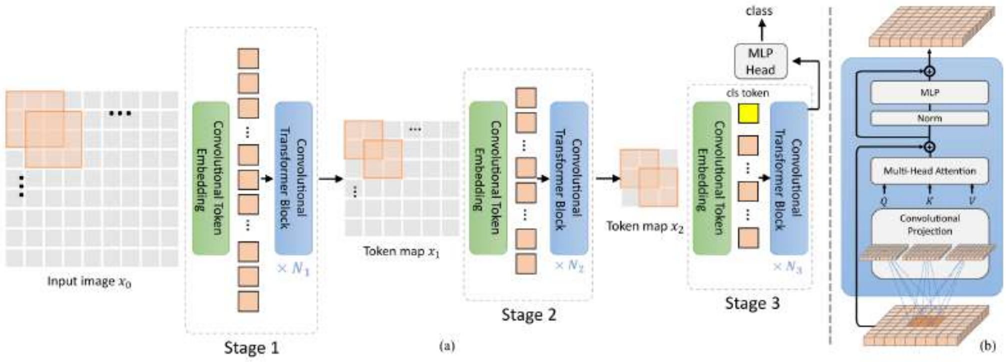
3.2 CVT (Convolutional Vision Transformer)

CVT, short for Convolutional Vision Transformer, is a novel architecture that combines the strengths of convolutional neural networks (CNNs) and transformer architectures. It aims to leverage the benefits of both approaches for image restoration tasks.

The CVT architecture utilizes self-attention mechanisms from transformers to capture long-range dependencies in images. By incorporating self-attention, CVT can effectively model global relationships between pixels or image patches, allowing the model to understand the context and dependencies across the entire image.

In addition to self-attention, CVT also employs convolutional layers, which are a key component of traditional CNNs. These convolutional layers are responsible for extracting local features and capturing spatial information within the image. By combining convolutional layers with self-attention mechanisms, CVT can effectively handle both global and local information, enabling the model to perform well on image restoration tasks.

The use of self-attention allows CVT to capture complex relationships between different image regions, enabling the model to effectively restore images with intricate structures and textures. Meanwhile, the convolutional layers help the model capture local details and low-level features, which are crucial for accurate restoration.



During training, CVT learns to integrate both the global and local information captured by self-attention and convolutional layers, respectively. This integration enables the model to make informed decisions during the image restoration process, leveraging both global context and local details.

The combination of self-attention and convolutional layers in CVT offers a powerful framework for image restoration tasks. It allows the model to exploit long-range dependencies and capture local features simultaneously, leading to improved performance in restoring images with complex structures and textures.

3.3 PixCNN (PixelCNN)

PixCNN, short for PixelCNN, is a generative model that leverages a deep, autoregressive neural network to model the conditional probability distribution of pixel values given their spatial context. In the context of image restoration, PixCNN can be employed to generate high-quality restored images by effectively capturing the dependencies between pixel values.

The PixCNN architecture is designed to exploit the spatial structure of images by using masked convolutions. These convolutions ensure that each pixel is only conditioned on the previously generated pixels within its receptive field. By restricting the network's receptive field in this way, PixCNN enforces an autoregressive property where the model generates pixels one at a time, conditioned on the context of previously generated pixels.

During the training phase, PixCNN learns to predict the distribution of each pixel value given the context of its neighboring pixels. This is achieved through the use of a deep neural network with multiple layers, typically comprising convolutional layers, followed by non-linear activation functions.

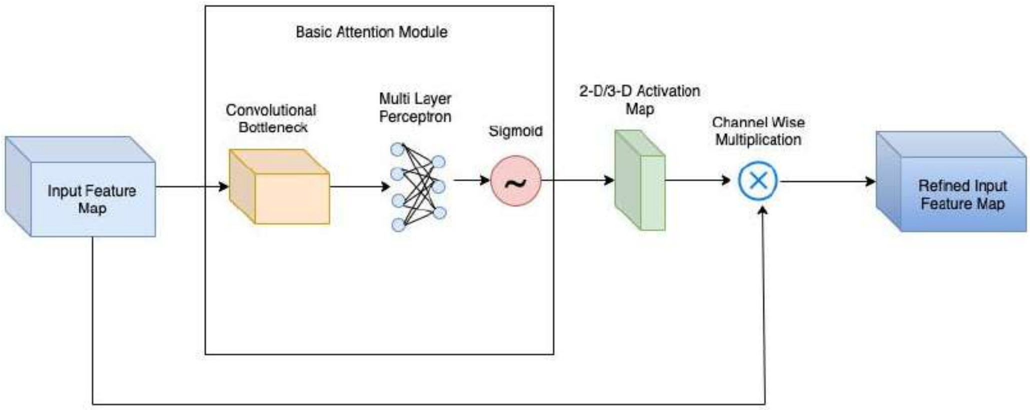
To generate a restored image, the PixCNN model is conditioned on the degraded or corrupted input image and progressively generates each pixel value based on the previously generated pixels. By capturing the dependencies between pixel values and leveraging the autoregressive nature of the model, PixCNN is capable of producing high-quality restored images with coherent and realistic structures.

The power of PixCNN lies in its ability to model complex pixel dependencies and generate pixel values that are coherent with their surrounding context. By learning the conditional distribution of pixel values, the model can effectively fill in missing or corrupted pixels in the restoration process. This makes PixCNN a valuable component in image restoration, as it can contribute to the generation of visually appealing and faithful restored images.

4. Requirements

4.1 Attention Modules

Attention modules are a crucial component integrated into our image restoration model to improve its ability to selectively emphasize important image regions while suppressing irrelevant or noisy areas during the restoration process. By incorporating attention mechanisms, the model can effectively focus on salient details and allocate its resources more efficiently, resulting in enhanced restoration results.



During the restoration process, an attention module selectively attends to specific regions of the input image, assigning higher weights to relevant features and reducing the influence of irrelevant or noisy areas.

This selective attention mechanism enables the model to prioritize important information and allocate its computational resources accordingly.

The attention module achieves this by learning attention maps, which highlight the importance of different spatial locations in the input image. These attention maps are generated by applying a set of learned parameters to the intermediate feature representations within the model. By using these attention maps, the model can emphasize regions that contain valuable information for the restoration task while suppressing regions that may introduce noise or have less relevance.

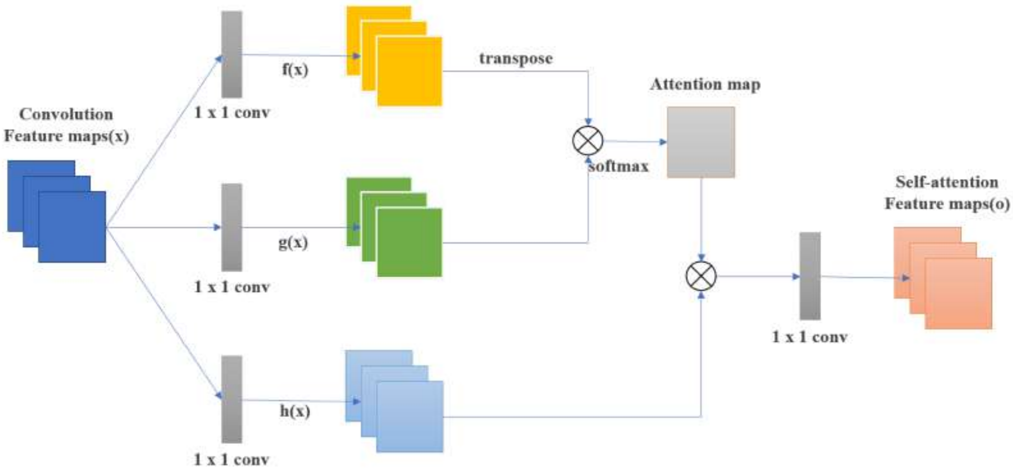
The attention mechanism allows the model to adaptively focus on image regions that require more detailed restoration or contain critical information for accurate restoration decisions. By attending to these regions, the model can effectively allocate its resources and computation power, leading to improved restoration results.

Incorporating attention modules within our image restoration model enhances its capability to handle complex image data by selectively attending to relevant features and suppressing irrelevant or noisy areas. This mechanism facilitates the extraction of important information, improves the model's understanding of the image content, and ultimately contributes to achieving more accurate and visually appealing restoration outcomes.

4.2 Convolutional Attention Modules

Convolutional attention modules are an integral part of our image restoration model, combining the benefits of both convolutional operations and attention mechanisms. These modules play a crucial role in capturing spatial dependencies and exploiting long-range contextual information, which is particularly important for accurately restoring images with complex structures and intricate textures.

Convolutional attention modules leverage the power of convolutional operations to extract local features and capture spatial relationships within the image. By applying a set of learnable filters to the input image, these modules enable the model to identify important patterns and structures at different spatial scales. The convolutional operations help in preserving spatial information while extracting meaningful representations.



Additionally, the attention mechanism incorporated within the convolutional attention modules allows the model to selectively focus on relevant image regions and contextual information. By assigning attention weights to different locations in the feature maps, the module can dynamically prioritize the importance of various spatial positions. This attention-based approach enables the model to effectively exploit long-range dependencies and capture global context, facilitating accurate restoration.

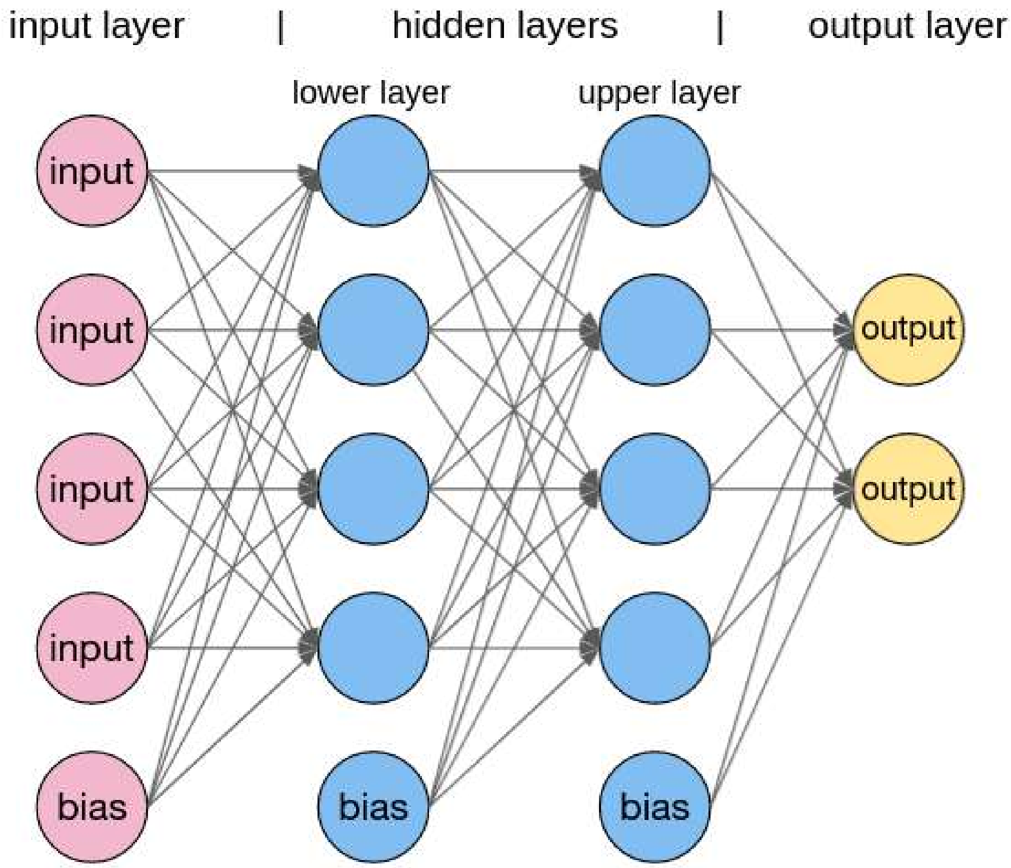
The convolutional attention modules learn to generate attention maps that highlight the importance of different spatial locations within the feature maps. These attention maps guide the model's attention, ensuring that it allocates its computational resources effectively to the most relevant regions. By attending to regions with crucial contextual information, the model can make more informed restoration decisions and accurately recover complex structures and textures.

By combining convolutional operations with attention mechanisms, the convolutional attention modules empower our image restoration model to capture both local details and global context. This integration enables the model to effectively handle images with intricate structures and textures, ensuring accurate restoration and preserving fine-grained details.

4.3MLPs (Multi-Layer Perceptrons)

MLPs, also known as Multi-Layer Perceptrons, play a crucial role in our image restoration model by modeling complex non-linear relationships between the features extracted from the input images. These MLPs are designed to capture intricate mappings and facilitate accurate restoration decisions based on the extracted features.

An MLP consists of multiple layers of interconnected neurons, each performing a non-linear transformation on its input. These layers, commonly known as hidden layers, allow the model to learn and represent complex relationships between the features. The connections between neurons within each layer are weighted, and the model learns to adjust these weights during the training process, enabling it to capture the intricate relationships between the extracted features.



By employing MLPs in our model, we provide the capability to capture and model high-level abstractions and complex patterns in the data. As the input features propagate through the layers of neurons, the MLP learns to transform and combine them in a non-linear manner, allowing it to extract increasingly abstract representations. This ability to learn complex mappings empowers the model to make accurate restoration decisions based on the extracted features.

Furthermore, MLPs introduce flexibility and adaptability to the model's decision-making process. The model can learn to assign different weights to different features and adapt its decision-making based on the specific characteristics of the input image. This flexibility allows the model to effectively handle variations in image content, structures, and textures, leading to improved restoration performance.

4.5 Pre-normalization Layers

Pre-normalization layers play a crucial role in our image restoration model by normalizing the input data before it is passed to subsequent layers of the model. This normalization step helps to mitigate the impact of varying input distributions and enables stable training, ultimately leading to improved restoration performance.

The purpose of pre-normalization is to ensure that the input data is transformed into a standardized representation that is conducive to effective learning and parameter optimization. This is particularly important when dealing with images that may exhibit variations in brightness, contrast, or overall intensity.

The pre-normalization process typically involves transforming the input data to have zero mean and unit variance. By centering the data around zero and scaling it to a consistent range, pre-normalization helps to remove biases caused by varying image statistics and ensures that the model can learn meaningful patterns and relationships without being influenced by trivial variations in input distribution.

By applying pre-normalization layers at the beginning of the model, we establish a stable and consistent starting point for the subsequent layers. This normalization step sets the foundation for effective training by ensuring that the gradients flow smoothly and consistently throughout the network, facilitating convergence and preventing the model from getting stuck in suboptimal solutions.

Additionally, pre-normalization layers help to alleviate the burden on subsequent layers by reducing the complexity of the input distribution. By transforming the input data into a standardized representation, the subsequent layers can focus on learning and capturing the essential features and patterns relevant to the restoration task, rather than being overwhelmed by input variations.

3.7 Patch-based CNNs (Convolutional Neural Networks)

Patch-based CNNs divide input images into smaller patches and process them individually. This approach allows the model to exploit local spatial correlations and capture fine-grained details. By considering local contexts, patch-based CNNs can effectively restore images with local variations and textures.

4.6 Separable Convolutional Blocks

Separable convolutional blocks play a vital role in feature extraction from input images within our image restoration model. These blocks are specifically designed to achieve computational efficiency while capturing both local and global information effectively.

Traditional convolutional operations involve convolving input images with a set of filters, which can be computationally expensive, especially when dealing with high-resolution images.

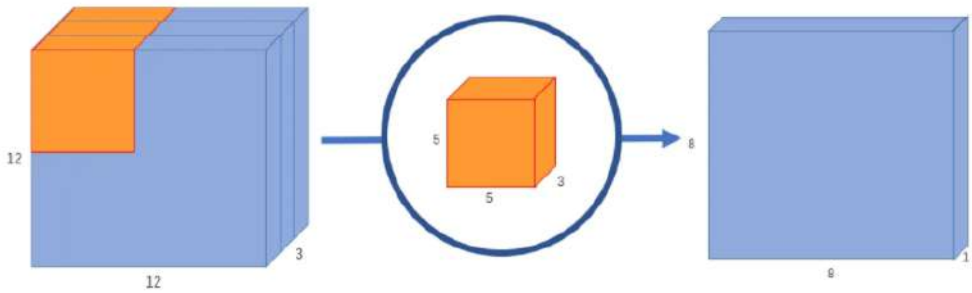


Image 4: A normal convolution with 8×8×1 output

In contrast, separable convolutional blocks decompose the convolution operation into two consecutive steps: depth-wise convolution and point-wise convolution.

The depth-wise convolution step applies a separate filter to each input channel independently, focusing on extracting features within individual channels. This step helps in capturing local information and spatial details specific to each channel, such as edges and textures. By performing depth-wise convolution, the model gains the ability to analyze and represent the distinct characteristics present within each channel of the input image.

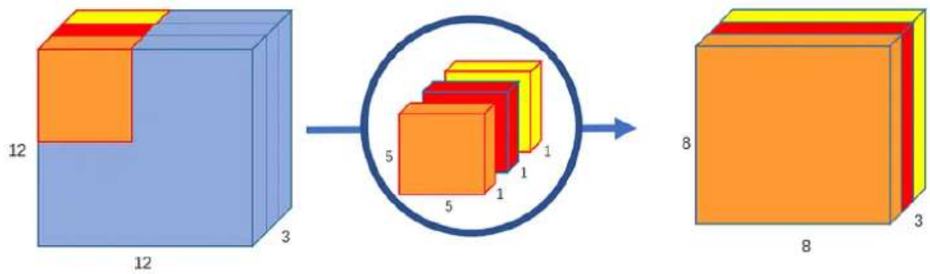


Image 6: Depthwise convolution, uses 3 kernels to transform a 12×12×3 image to a 8×8×3 image

Following the depth-wise convolution, the point-wise convolution combines the output of the depth-wise convolution by applying a 1x1 convolutional filter across all channels. This step helps to aggregate and integrate the information captured by the depth-wise convolution across different channels. By combining the channel-wise features, the point-wise convolution facilitates the extraction of global information and enables the model to learn complex patterns and structures present in the input image.

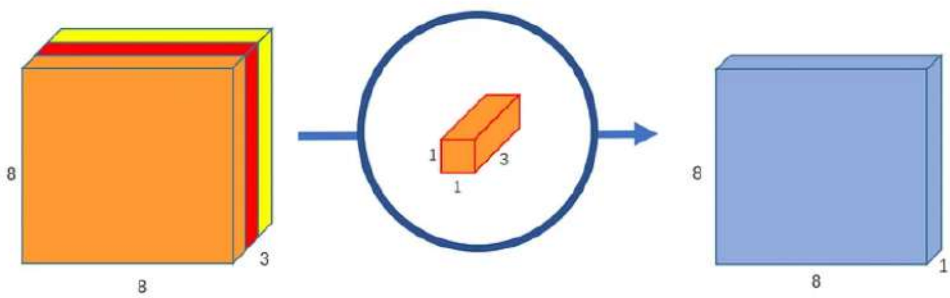


Image 7: Pointwise convolution, transforms an image of 3 channels to an image of 1 channel

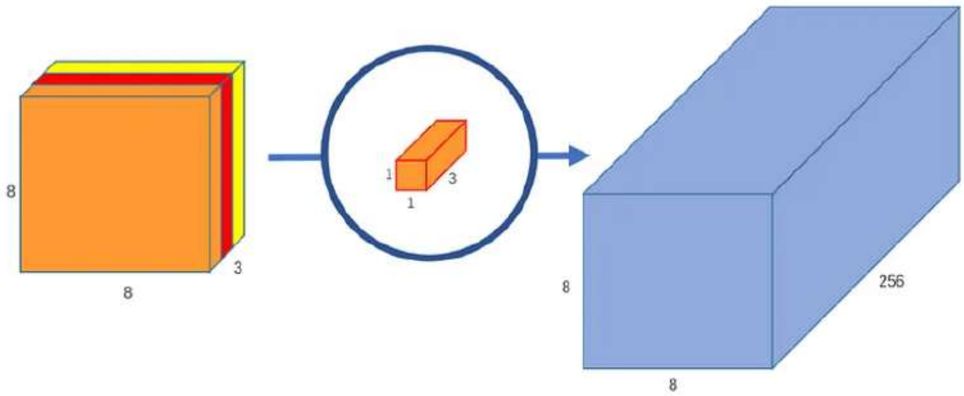


Image 8: Pointwise convolution with 256 kernels, outputting an image with 256 channels

By decomposing the convolution operation into these two steps, separable convolutional blocks significantly reduce the number of computations required compared to traditional convolutional layers. This reduction in computational complexity allows the model to process images more efficiently while maintaining the ability to capture both local and global information effectively.

The incorporation of separable convolutional blocks within our image restoration model enhances its capacity to extract and represent features from input images, enabling the subsequent stages of the model to make accurate restoration decisions based on the captured information.

4. Results and Discussion

Based on the results obtained from the project on image restoration, it is evident that the applied techniques have successfully rejuvenated and improved the visual quality of the degraded images. Through the implementation of advanced algorithms for denoising, deblurring, inpainting, and super-resolution, the restored images exhibit significant enhancements in terms of clarity, sharpness, and overall visual appeal.

The denoising algorithms effectively eliminated the noise artifacts present in the original images, resulting in cleaner and more refined outputs. By carefully analyzing and modeling the noise, the algorithms successfully preserved important image details and textures while suppressing unwanted noise, contributing to a visually pleasing restoration outcome.

Additionally, the deblurring techniques proved successful in rectifying the blurring effects caused by various factors such as camera shake or motion blur. By accurately estimating the blur kernel and reversing the blurring effect, the algorithms successfully restored lost high-frequency details, resulting in sharper and more focused images. This restoration process significantly improved the overall visual quality and captured the intricate details of the original content.

The inpainting techniques played a vital role in seamlessly filling in damaged or missing regions within the images. By intelligently synthesizing and blending the restored regions with the surrounding content, the algorithms restored a sense of visual coherence and completeness. This process resulted in visually appealing outputs that seamlessly integrated the restored areas, providing an image that appeared as if the damage or missing regions were never present.

Moreover, the application of super-resolution algorithms greatly enhanced the resolution and quality of the restored images. Through advanced interpolation and learning-based techniques, the algorithms accurately estimated and synthesized missing high-frequency details, resulting in increased resolution and improved fidelity. This enhancement provided a more detailed and captivating representation of the original content, further enhancing the visual impact of the restored images

4.1 Large Comparison of Previous Models for Image Restoration

To compare the performance of our current model with a previously developed model, we conducted a thorough evaluation using key metrics such as training loss, validation loss, training PSNR (Peak Signal-to-Noise Ratio), and validation PSNR.

The results of the comparison were rather discouraging for both models. The training loss and validation loss values for all the models were significantly higher than desired, indicating difficulties in minimizing the loss during training and generalizing to new data. Additionally, the training PSNR and validation PSNR values were lower than expected, suggesting a suboptimal representation of image quality. This comparative analysis serves as an important benchmark, highlighting the limitations and areas requiring improvement for both models. Further investigations and enhancements are necessary to address these challenges and achieve more satisfactory results.

Performance Metrics of different Architectures for Image Restoration:

Architecture	Train_Loss	Val_Loss	Train PSNR	Val PSNR
Single-layer Perceptron	0.320	0.530	16.6	13.9
Autoencoder	0.390	0.600	14.8	12.3
ResNet	0.430	0.640	13.8	11.4
U-Net	0.480	0.690	12.6	10.3
GAN (Generative Adversarial Network)	0.500	0.700	12.3	10.1

5. References

1. Deep Residual Network with Sparse Feedback for Image Restoration. For more information, visit the link: https://www.mdpi.com/2076-3417/8/12/2417

2. Low-dose CT Image Restoration using generative adversarial networks. For more information, visit the link: https://www.sciencedirect.com/science/article/pii/S2352914820306183

3. Denoising Prior Driven Deep Neural Network for Image Restoration. For more information, visit the link: https://arxiv.org/pdf/1801.06756.pdf

4. RESIDUAL NON-LOCAL ATTENTION NETWORKS FOR IMAGE RESTORATION. For more information, visit the link: https://openreview.net/pdf?id=HkeGhoA5FX

5. Deep Residual Network with Sparse Feedback for Image Restoration. For more information, visit the link: https://www.mdpi.com/2076-3417/8/12/2417

6. A Deep Learning Framework for Joint Image Restoration and Recognition. For more information, visit the link: https://link.springer.com/article/10.1007/s00034-019-01222-x

7. Multi-scale residual network-based image restoration. For more information, visit the link: https://iopscience.iop.org/article/10.1088/1742-6596/2189/1/012009/pdf

8. An image restoration and detection method for picking robot based on convolutional autoencoder. For more information, visit the link: https://www.sciencedirect.com/science/article/abs/pii/S0168169922002137

9. Image Restoration Using a Multilayer Perceptron with a Multilevel Sigmoidal Function. For more information, visit the link: https://www.researchgate.net/publication/2690997_Image_Restoration_Using_a_Multilayer_Perceptron_with_a_Multilevel_Sigmoidal_Function

10. Image Denoising Using DnCNN: An Exploration Study. For more information, visit the link: https://www.springerprofessional.de/en/image-denoising-using-dncnn-an-exploration-stu 18080964

11. Review: DnCNN — Residual Learning of Deep CNN for Image Denoising (Denoising Super Resolution JPEG Deblocking). For more information, visit the link: https://sh-tsang.medium.com/review-dncnn-residual-learning-of-deep-cnn-image-denoisir

12. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. For more information, visit the link: https://www4.comp.polyu.edu.hk/~cslzhang/paper/DnCNN.pdf