**Image Restoration Using Deep Learning Techniques**

***Dissertation submitted to the***

***University College of Engineering (A), Osmania University.***

***In Partial fulfilment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

In

**COMPUTER SCIENCE AND ENGINEERING**

**Submitted by**

**Mogadampalli Mani datt (100521733036)**

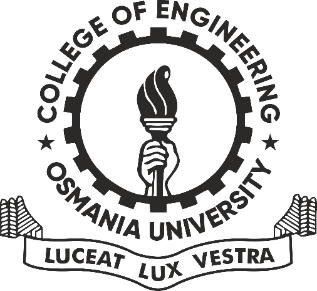
**Puli Venu (100521733045)**

**Kspsvln Siddardha Kumar Kavuri (100521733058)**

***Under the Guidance of***

***Professor P.V Sudha***

**Dept. of CSE, UCEOU.**

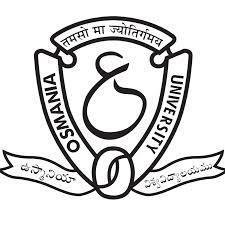


**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**UNIVERSITY COLLEGE OF ENGINEERING (A),**

**OSMANIA UNIVERSITY HYDERABAD, TS-500007**

**MAY - 2025**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**UNIVERSITY COLLEGE OF ENGINEERING (A),**

**OSMANIA UNIVERSITY**

**CERTIFICATE**

This is to certify the bonafide work of **Mogadampalli Mani datt, Puli Venu, Kspsvln Siddardha Kumar Kavuri** bearing roll numbers **100521733036,100521733045,**

**100521733058** respectively, for their major project, in partial fulfilment of Bachelor of Engineering Degree, offered by Department of Computer Science and Engineering, University College of Engineering, Osmania University.

**Project Guide Head of the Department**

Prof. P.V. Sudha Prof. P.V. Sudha

Dept. of CSE, UCEOU Dept. of CSE, UCEOU

**STUDENT DECLARATION**

We declare that the wok reported in the project report entitled “Image Restoration Using Deep Learning Techniques ” submitted by **Mogadampalli Mani datt, Puli Venu, Kspsvln Siddardha Kumar Kavuri** is a record of the work done by us in the DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY COLLEGE OF ENGINEERING . No part of the report is copied from books/journals/internet and wherever referred, the same has been duly acknowledged in the text . The reported data is based on the work done entirely by us and not copied from any other source or submitted to any other Institute or University for the award of a degree or diploma.

**SIGNATURES**

**Mogadampalli Mani Datt :**

**Puli Venu :**

**KSPSVLN Siddardha Kumar Kavuri :**

**ACKNOWLEGMENT**

It is our privilege and pleasure to express our profound sense of respect, gratitude, and indebtedness to our guide, Prof. P. V. Sudha, Head, Department of Computer Science and Engineering, UCEOU, for her constant inspiration, valuable guidance, insightful discussions, constructive criticisms, and unwavering encouragement throughout this dissertation work. We are also deeply thankful for her support from the Department and for providing access to the necessary resources, which played a crucial role in the successful completion of this project.

**ABSTRACT**

In an era dominated by visual data, the quality and clarity of images play a pivotal role across various domains, from medical diagnostics to autonomous driving and surveillance systems. Our project focuses on revolutionizing the field of image restoration by leveraging advanced artificial intelligence to recover degraded, noisy, or blurred images with remarkable precision and efficiency. Through the integration of cutting-edge deep learning architectures, including Convolutional Neural Networks, Transformers, and Generative Adversarial Networks, our system addresses key challenges such as noise reduction, blur removal, motion deblurring, and low-light enhancement.

By automating and enhancing the image restoration pipeline, we aim to significantly reduce manual post-processing, improve real-time applicability, and enable high-fidelity image recovery in challenging scenarios. The use of benchmark datasets such as SIDD, GoPro, and HIDE ensures that our approach is rigorously evaluated and adaptable to real-world conditions. Furthermore, by embedding our system within broader computer vision workflows, we pave the way for intelligent applications in security, healthcare, and mobile imaging.

Ultimately, this project represents a leap forward in the intelligent processing of visual data, showcasing how AI can elevate the quality of information extracted from degraded inputs and support faster, more informed decision-making in critical visual environments.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **CONTENT** | **PAGE NO.** |
| **1.** | Chapter 1: Introduction   * 1. Introduction   2. Aim   3. Deep Learning   4. Convolutional Neural Network   5. Transformer Learning   6. UNET Architecture | 1  1  2  2  3  5  7 |
| **2.** | Chapter 2 : Literature Survey | 9 |
| **3.** | Chapter 3 : Existing Methods | 13 |
| **4.** | Chapter 4: Proposed Method   * 1. Proposed Approach      1. Sparse Transformer      2. DGUNET-Based deblurring | 18  18  18  20 |
| **5.** | Chapter 5: Results | 22 |
| **6.** | Chapter 6 : Performance and Score Evaluation | 30 |
| **7.** | Chapter 7: Conclusion and Future Scope   * 1. Conclusion   2. Future Scope | 31  31  32 |
| **8.** | References | 33 |

**CHAPTER I**

1.INTRODUCTION

**1.1 Introduction**

Images serve as a powerful medium of information in various domains, including medical diagnostics, surveillance, historical documentation, and astronomy. However, real-world images are often degraded due to noise, motion blur, low resolution, occlusions, or environmental damage. This degradation results in loss of critical information, which can hinder accurate interpretation and decision-making. Image restoration plays a crucial role in reviving visual information from such degraded inputs and is thus an essential area in the field of computer vision.

Traditional image restoration techniques based on mathematical models and handcrafted features often fall short when handling complex noise patterns or severe degradation. In recent years, deep learning has revolutionized image processing, offering powerful models that can learn intricate mappings between degraded and clean images. These data-driven approaches—particularly Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Transformers—have demonstrated significant success in various restoration tasks such as denoising, deblurring, super-resolution, and colorization.

A surge in publicly available datasets, such as the GoPro dataset for deblurring and other benchmark datasets for denoising and super-resolution, has further accelerated the development of robust restoration models. The advancement in GPU hardware and open-source deep learning frameworks like PyTorch has also contributed to rapid experimentation and deployment of these models in real-world applications.

This project presents a comprehensive image restoration framework that integrates CNN and Transformer architectures to effectively extract and reconstruct image features. The proposed method aims to restore fine details, enhance visual quality, and generate high-resolution outputs from severely degraded inputs. Applications of this work span multiple industries, including medical imaging, digital forensics, historical preservation, and consumer photography.

With the growing demand for high-quality visual data, especially in fields like autonomous driving, telemedicine, satellite imaging, and augmented reality (AR), the ability to restore degraded images has become more critical than ever. Restoring such images not only improves aesthetic quality but also enhances the performance of downstream tasks like object detection, segmentation, and recognition. For instance, in surveillance systems, improving image clarity can be vital for identifying faces or license plates; in medical diagnostics, restoring noisy scans can lead to more accurate disease detection. This project leverages recent advancements in deep learning-based restoration techniques to develop a scalable and efficient model capable of handling various types of image degradation. The end goal is to build a generalized restoration pipeline that maintains structural integrity and visual realism across diverse scenarios.

* 1. **AIM**

The aim of this project is to leverage state-of-the-art deep learning techniques to automate and enhance the process of image restoration, with a focus on recovering lost visual information from degraded, blurred, or low-resolution images. By integrating convolutional and transformer-based neural networks, the goal is to develop a generalized restoration model capable of reviving visual quality across diverse image domains such as medical imaging, historical preservation, surveillance, and astronomy. Specific objectives of this approach include:

* **Enhanced Visual Quality**: Deploy deep learning algorithms to restore details, reduce noise, and correct distortions, significantly improving the perceptual quality of images.
* **Automation and Scalability**: Build an automated system that can process large volumes of degraded images without human intervention, making it scalable across industries.
* **Real-Time Restoration**: Enable real-time processing of video frames and images for applications like surveillance monitoring and live medical imaging support.
* **Cross-Domain Applicability**: Ensure the restoration model is robust and adaptable to various domains including satellite imagery, consumer photography, and old archival content.
* **Resource Optimization**: Reduce reliance on manual retouching or traditional filters by providing an intelligent, learning-based solution that evolves with time.
* **Ethical Use and Authenticity**: Incorporate features to distinguish between authentically restored and artificially manipulated images, supporting ethical AI use and digital integrity.

By achieving these objectives, the project aims to offer a comprehensive AI-powered solution that revolutionizes how degraded images are handled and restored, ultimately contributing to enhanced visual communication and decision-making in critical fields.

* 1. **DEEP LEARNING**

Deep learning is a branch of artificial intelligence (AI) that mimics the way the human brain works to process data and make decisions. At its core, deep learning involves training algorithms known as neural networks to recognize patterns and extract meaningful information from large datasets. These neural networks are composed of interconnected layers of nodes, each performing simple mathematical operations on the data they receive.

One of the key strengths of deep learning is its ability to automatically learn features from raw data, eliminating the need for manual feature engineering. This makes it particularly well-suited for tasks such as image and speech recognition, natural language processing, and medical diagnosis, where the input data may be complex and high-dimensional.

Deep learning has achieved remarkable success in a wide range of applications, from powering virtual assistants like Siri and Alexa to enabling self-driving cars and improving healthcare outcomes. Its ability to process vast amounts of data and extract valuable insights has revolutionized industries and transformed the way we interact with technology.

Despite its power and versatility, deep learning does have some limitations. It requires large amounts of labelled data for training, which can be time-consuming and expensive to collect. Additionally, deep learning models can be complex and computationally intensive, requiring significant computational resources for training and inference.

Overall, deep learning holds tremendous potential for solving complex problems and driving innovation across various domains. As researchers continue to advance the field and develop more efficient algorithms and techniques, the impact of deep learning is expected to grow even further in the years to come.

* 1. **CONVOLUTIONAL NEURAL NETWORK:**

The architecture of a Convolutional Neural Network (CNN) is structured into three primary components: the input layer, the hidden layers, and the output layer.

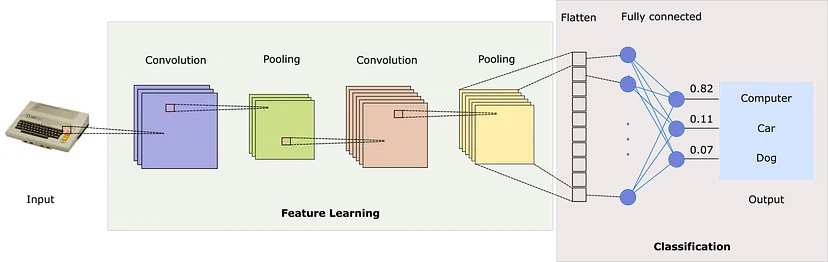


Fig 1.1 Architecture of Convolutional Neural Network

**1.Input Layer** : The input layer serves as the entry point for the network. It receives the input image data, which is typically in the form of a matrix of pixel values representing the image's features.

**2.Hidden Layers** :The hidden layers are where the core processing of the CNN occurs. They consist of multiple convolutional and pooling layers.

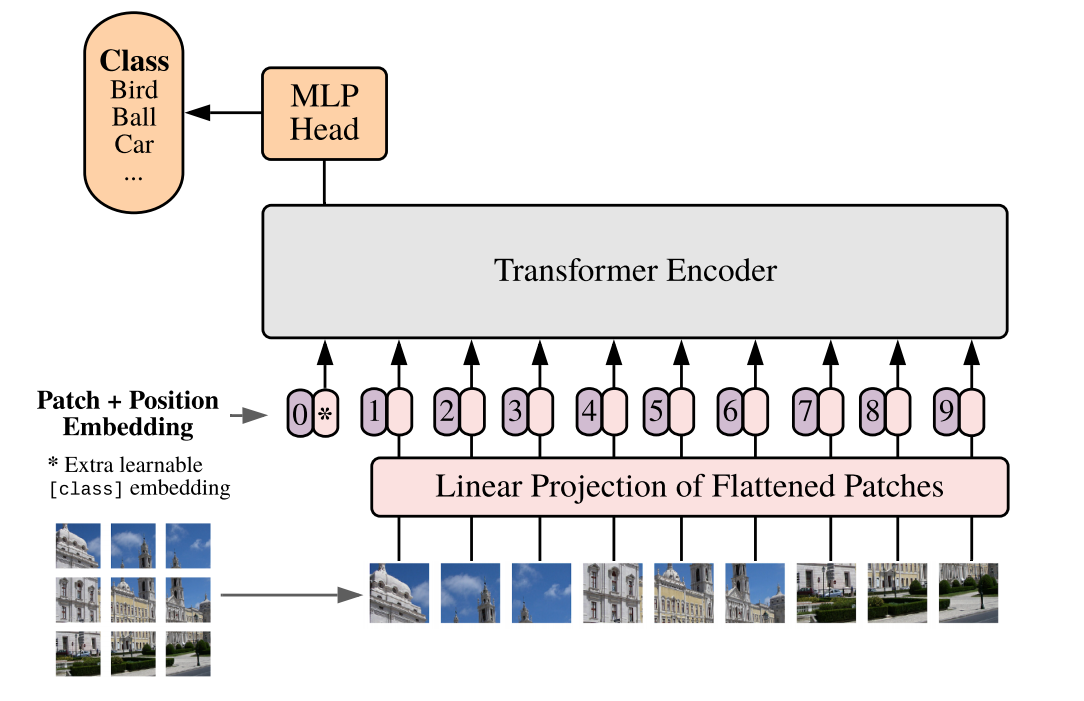
* **Convolutional Layers** : These layers apply a set of learnable filters (also called kernels) to the input image. Each filter detects specific patterns or features, such as edges or textures, by performing convolution operations. This helps the network to extract hierarchical representations of the input data.
* **Pooling Layers** : After each convolutional layer, a pooling layer is often applied to down sample the feature maps generated by the convolutional layers. Pooling helps to reduce the spatial dimensions of the feature maps while retaining important information, thereby decreasing computational complexity and controlling overfitting.

**3.Output Layer** : The output layer provides the final results of the network's processing. It typically consists of one or more fully connected layers that perform classification or regression tasks based on the features extracted by the hidden layers. For classification tasks, the output layer produces the predicted class label or probability scores for each class.

The configuration of the hidden layers is crucial in determining the CNN's performance. Parameters such as the number of hidden layers, the number of filters in each convolutional layer, and the choice of pooling strategies can be adjusted to optimize the network's accuracy and efficiency.

**1.5 Transformer Learning**

Transformer Learning represents a major advancement in deep learning, originally introduced in the domain of Natural Language Processing (NLP) and now increasingly adopted in computer vision tasks such as image classification, segmentation, and restoration. Unlike traditional CNNs that operate using fixed receptive fields, Transformers rely on the **self-attention mechanism** to model global relationships between pixels or patches in an image. This capability enables them to capture long-range dependencies and complex structures that are difficult for CNNs to learn.



**Fig 1.4 Transformer Architecture in Vision Tasks**

The core component of a transformer is the **multi-head self-attention** module, which allows the model to weigh the importance of different regions of an image when making predictions. This attention mechanism is suppor-ted by **positional encoding**, which injects spatial information into the model, as unlike CNNs, Transformers are not inherently aware of spatial structure.

In vision applications, a common adaptation is the **Vision Transformer (ViT)**, which splits an input image into a sequence of patches (e.g., 16×16 pixels), flattens them, and treats them similarly to word tokens in NLP. These patches are then linearly embedded and passed through a series of transformer encoder layers. Each layer consists of multi-head self-attention, feed-forward neural networks, normalization layers, and residual connections, making the architecture both deep and capable of learning fine-grained relationships.

**Attention Mechanism**

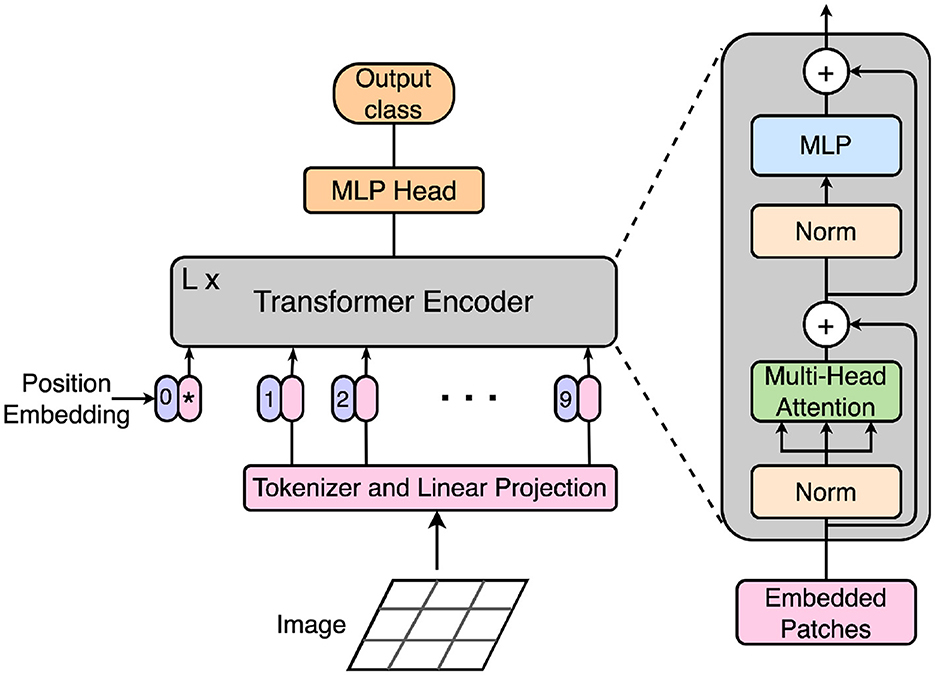
The self-attention mechanism computes a weighted average of all positions in the image, enabling the model to focus on the most informative parts. This mechanism is defined by three vectors:

* **Query (Q)**
* **Key (K)**
* **Value (V)**

The output is calculated as:

Attention(Q, K, V) = softmax(QKᵀ / √dₖ) \* V

Where dₖ is the dimensionality of the key vectors. This allows the model to compare and relate every image patch to every other patch, resulting in a better understanding of global features.



**Fig 1.5 Self-Attention in Transformer**

In modern image restoration systems, transformers are often combined with CNNs in hybrid architectures—CNNs for local feature extraction and transformers for capturing global context—resulting in powerful models like **Restormer**, **SwinIR**, and **PromptIR**.

**1)Encoder-Decoder Structure (U-net Back Bone):**

The encoder-decoder structure forms the core of the image restoration model, functioning as a U-Net-based architecture. The encoder is responsible for extracting key features from the input image by applying multiple convolutional layers while progressively reducing its spatial dimensions. It employs techniques such as gated convolutions to selectively retain important information and suppress noise. Additionally, the dual-branch structure allows the model to capture both low-frequency and high-frequency details, ensuring the preservation of textures and edges.

At the bottleneck, deep convolutional layers further refine the feature representation, incorporating multi-scale feature fusion to capture information at different scales. The supervised attention mechanism (SAM) helps the model focus on essential regions, enhancing its ability to restore degraded images effectively. This compressed yet meaningful representation serves as a bridge between the encoding and decoding stages.



The decoder takes the learned feature representation and reconstructs a high-quality image by progressively upsampling and refining the features. It utilizes transposed convolutions to restore the original image size while integrating skip connections from the encoder, which helps recover lost details. The application of gated convolutions in the decoder further improves image clarity by eliminating residual noise. Through the attention mechanism, the model emphasizes critical regions, ensuring a more accurate restoration process.

By leveraging U-Net’s architecture, the model effectively balances feature extraction and reconstruction, making it suitable for image restoration tasks. Skip connections prevent loss of important textures, while multi-scale fusion and attention mechanisms enhance the overall clarity and quality of the restored images. This approach ensures that fine details are preserved, making the restoration process efficient and reliable even with limited training data.

**CHAPTER II**

2.LITERATURE SURVEY

Several research studies have explored image restoration techniques using deep learning, with significant focus on architectures such as CNNs, GANs, and Transformers. These methods aim to enhance the quality of degraded images by reducing noise, blur, and resolution issues. Below are some key studies and models relevant to the advancement of this field:

**1)Liang et al. [1] (2021)**

**SwinIR**, a Swin Transformer-based architecture, was introduced for image restoration tasks including denoising and super-resolution. It employs a hierarchical design with shifted window attention to capture both local and global features efficiently. The model outperforms previous CNN-based methods by leveraging long-range dependencies while maintaining low computational cost. Evaluation on benchmark datasets showed superior PSNR and SSIM metrics, indicating high-quality restoration with preserved image structure.

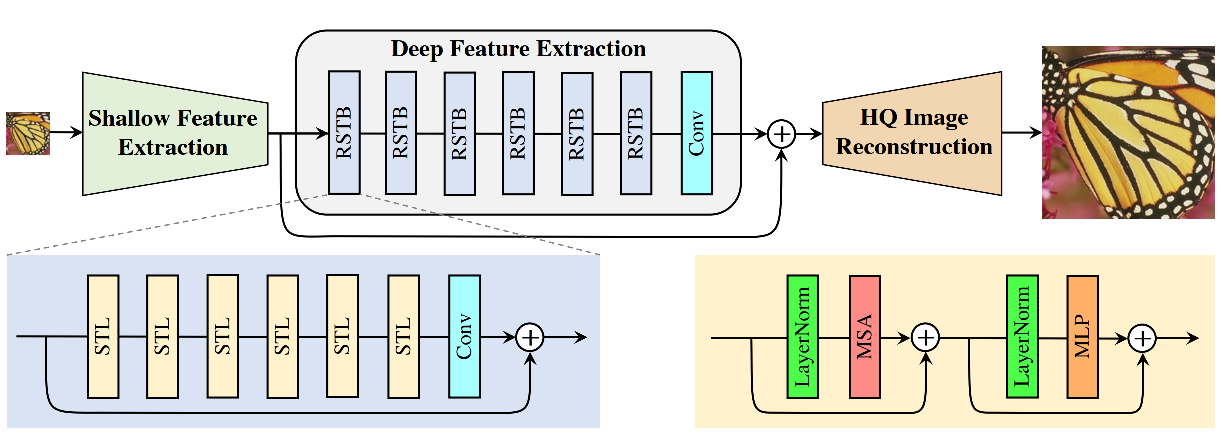


Fig Image Restoration Using Swin Transformer

**2)Zamir et al. [2] (2022)**

**Restormer** is a transformer-based architecture optimized for high-resolution image restoration. It introduces Multi-Dconv Head Transposed Attention (MDTA) and Gated-DFFN (GDFN) blocks for efficient feature representation and transformation. The network is lightweight yet capable of performing well on denoising, motion deblurring, and defocus correction tasks. The authors demonstrate state-of-the-art results on GoPro, REDS, and RealBlur datasets.

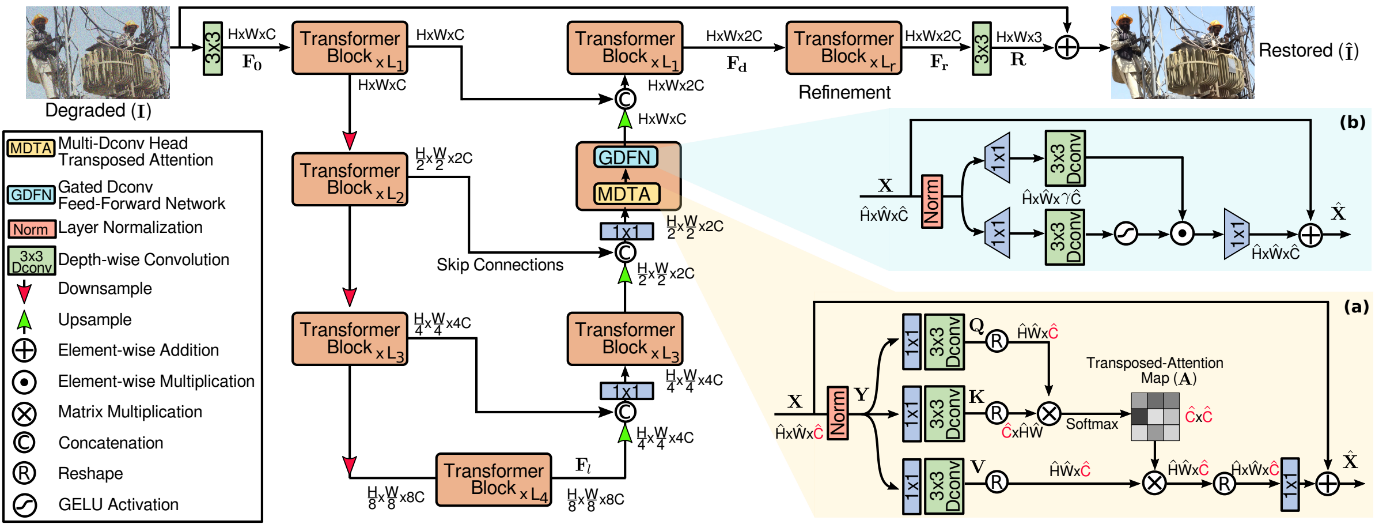


Fig Efficient transformer for High Resolution Image Restoration.

**3)Xia et al. [3] (2023)**

**DiffIR** investigates the use of diffusion models for image restoration. These models gradually remove noise from corrupted images using a reverse diffusion process guided by learned priors. Unlike conventional generative approaches, diffusion-based techniques yield highly realistic images with fine texture and detail recovery. This method proves particularly effective in challenging scenarios such as low-light and real-world denoising.



Fig Diffusion Model for Image Restoration

**4)Potlapalli et al. [4] (2023)**

**PromptIR** proposes an innovative solution that enables a single model to perform multiple restoration tasks based on user-defined prompts such as "denoise" or "deblur." The model utilizes shared transformer encoders and learns prompt embeddings, allowing it to dynamically adjust its behavior depending on the input instruction. This approach reduces the need for training task-specific models, promoting reusability and scalability.

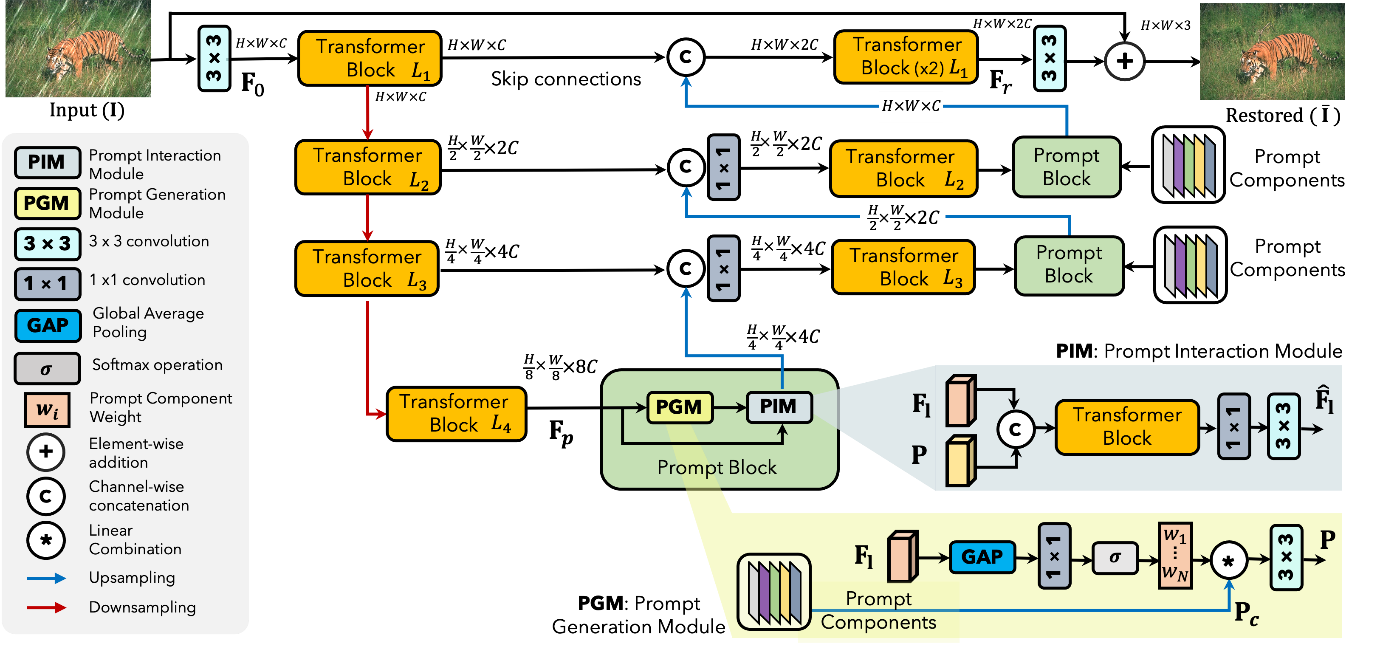


Fig Prompting for All-in-one blind Image Restoration

**5)Conde et al. [5] (2023)**

**InstructIR** builds upon PromptIR by integrating more flexible and natural language instructions to guide restoration. Users can specify detailed restoration needs (e.g., “remove motion blur and sharpen facial features”), and the model aligns image and language embeddings to achieve the desired output. This model demonstrates strong performance in interactive image enhancement systems and personalized image editing applications.

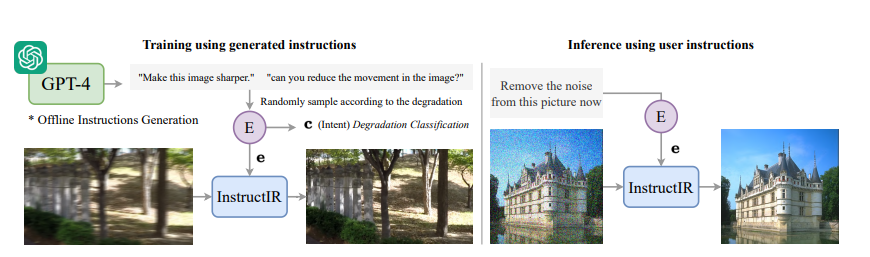


Fig Image Restoration using InstructIR

**6)Zhang et al. [6] (2022)**

**ELAN (Efficient Long-Range Attention Network)** combines shallow and deep feature extraction pipelines to restore high-quality images from heavily degraded inputs. The network uses 3x3 convolutions and global shortcut connections to boost feature learning efficiency. Results show significant improvement in super-resolution tasks, with better edge recovery and texture consistency compared to baseline CNNs.

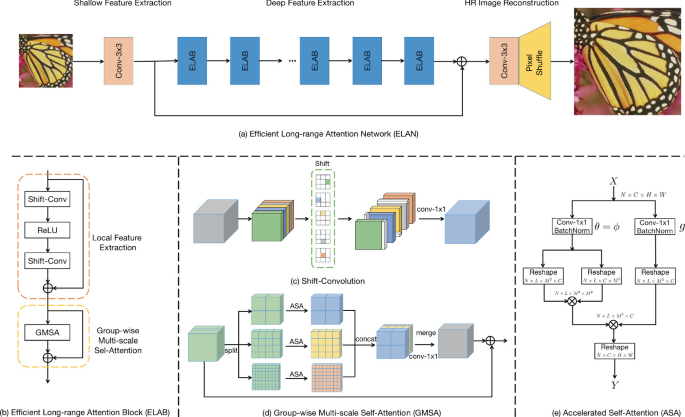


Fig Efficient long-range attention network (ELAN)

**7)Guo et al. [7] (2021)**

Investigated U-Net and its variants for segmenting and restoring damaged regions in medical and satellite imagery. The encoder-decoder structure with skip connections helps preserve spatial information, enabling accurate restoration of occluded and degraded areas. The method demonstrated superior performance in structured noise removal and partial image inpainting.

**8)Khan et al. [8] (2021)**

Demonstrated how GANs can be used for restoring extremely degraded images. Their study showed that adversarial training helps produce realistic textures and visual consistency in restored outputs. GANs are particularly effective in tasks like inpainting and low-resolution enhancement, where pixel-level accuracy is insufficient without perceptual guidance.

**CHAPTER III**

3.EXISTING METHODS

**3.1 Deblur GAN: Generative Adversarial Network for Image Deblurring**

**Deblur GAN** is a deep learning framework that leverages the power of Generative Adversarial Networks (GANs) to perform motion deblurring. It treats the task of deblurring as an image-to-image translation problem where the generator learns to transform a blurry image into its sharp counterpart.

**Key Components:**

1. **Generator Network**:  
   A Res Net-based architecture with residual blocks, responsible for producing a deblurred image from the input blurred image.
2. **Discriminator Network**:  
   A CNN-based binary classifier that learns to distinguish between real (ground truth sharp) images and generated (deblurred) ones.
3. **Perceptual Loss**:  
   Instead of using just pixel-wise loss, Deblur GAN incorporates perceptual loss using a pretrained VGG network, ensuring visually pleasing results by matching high-level features.

**Methodology:**

1. **Training Phase**:  
   The model is trained adversarially. The generator tries to fool the discriminator, while the discriminator learns to correctly identify real vs. generated images. Loss functions include:
   * Adversarial Loss
   * Content (L1) Loss
   * Perceptual Loss
2. **Inference Phase**:  
   A single forward pass through the generator restores the blurred input image into a sharp image.

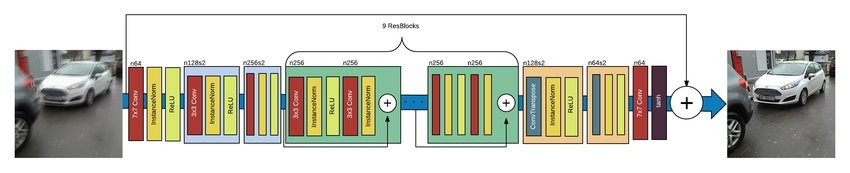


Fig Architecture of Deblur GAN generator network

**Advantages:**

* **Fast Inference**: Real-time performance on GPU due to efficient architecture.
* **High Visual Quality**: Generates realistic and sharp images.
* **No Paired Data Required** (DeblurGAN-v2): Unsupervised variants can work with unpaired data using cycle-consistency.

**Limitations:**

* **Sensitive to Training Stability**: GAN training is notoriously unstable.
* **Artifact Generation**: May produce artifacts in complex scenes if not carefully tuned.

**3.2 DiffIR: Diffusion Model for Image Restoration**

**DiffIR** is a diffusion-based image restoration method that models the data generation process as a **gradual denoising procedure**, learning to reverse this process for restoration. It is especially powerful in handling severe degradation.

**Key Components:**

1. **Forward Process (Diffusion)**:  
   Gaussian noise is progressively added to clean images in multiple steps until the image becomes nearly pure noise.
2. **Reverse Process (Denoising)**:  
   A neural network is trained to reverse this diffusion process by predicting the noise at each step and subtracting it iteratively.
3. **Conditioning Mechanism**:  
   DiffIR can condition the reverse process on the input degraded image, allowing flexible restoration of various degradation types (e.g., blur, noise, low-res).

**Methodology:**

1. **Training Phase**:  
   The model learns a denoising function for each diffusion step using supervised training on clean-degraded image pairs. Typically uses a **U-Net** backbone and loss functions like L2 or perceptual loss.
2. **Inference Phase**:  
   Starts from a noisy version of the degraded image and iteratively denoises it over hundreds or thousands of steps to obtain the final clean image.



Fig Diffusion Model for Image Restoration

**Advantages:**

* **High-Quality Restoration**: Restores fine textures and preserves structure better than GANs in many cases.
* **Theoretically Grounded**: Based on probabilistic modelling with controllable sampling.
* **Works Well on Real-World Data**: Effective even when the degradation is complex or unknown.

**Limitations:**

* **Slow Inference**: Requires many forward passes to complete the denoising steps.
* **High Computational Demand**: Training and inference are resource-intensive.
* **Long Training Time**: Needs significant data and training time to converge.

**3.3 DnCNN: Denoising Convolutional Neural Network**

**DnCNN** is one of the earliest and most influential deep learning models for image denoising. Introduced by Zhang et al., it outperforms traditional methods by learning a **residual mapping** from noisy images to clean ones. Rather than directly predicting the clean image, it learns to estimate the noise, which is then subtracted from the input image to obtain the denoised result.

**Key Components:**

1. **Convolutional Layers**:  
   A deep stack of convolutional layers with small kernels (typically 3x3), ReLU activations, and batch normalization.
2. **Residual Learning Strategy**:  
   Instead of predicting the clean image, the network predicts the noise (residual), which is subtracted from the noisy input.
3. **Loss Function**:  
   The model uses **Mean Squared Error (MSE)** loss between the predicted noise and the actual noise in training pairs.

**Methodology:**

1. **Training Phase**:
   * Input: Noisy image
   * Target: Noise image (clean - noisy)
   * The network learns to minimize the MSE between predicted and ground-truth noise.
2. **Inference Phase**:
   * The network predicts the noise from a given noisy image.
   * The denoised output is obtained by subtracting the predicted noise from the input image.

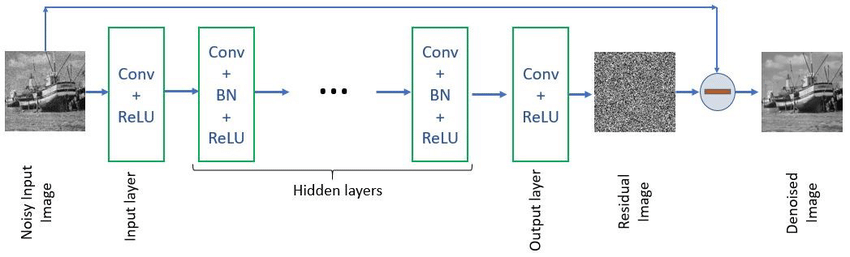


Fig DnCNN Network Architecture

**Advantages:**

* **Simple and Efficient**: Uses only standard convolutional layers; easy to implement and fast to train.
* **Effective Across Noise Levels**: Can be trained for both blind and non-blind Gaussian noise levels.
* **Generalizable**: With minor modifications, it has been applied to other tasks like super-resolution and JPEG artifact reduction.

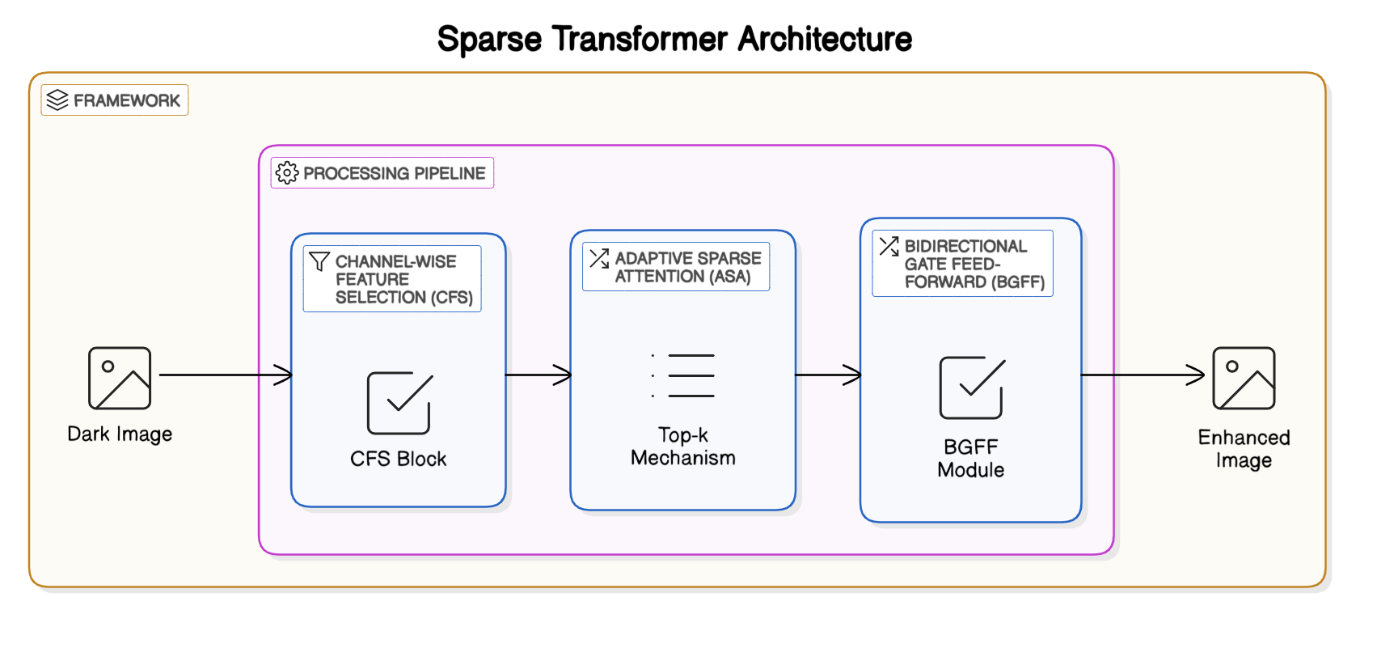
**Limitations:**

* **Local Receptive Field**: Lacks the ability to model long-range dependencies in images, limiting performance on complex degradations.
* **Not Adaptive**: Requires separate training for different types or levels of noise unless trained in a blind fashion.
* **Limited for Real-World Noise**: Performance may drop on real-world noise, which often deviates from Gaussian assumptions.

**CHAPTER IV**

4. PROPOSED METHOD

This chapter outlines two advanced deep learning-based image restoration techniques explored in this study: one using a **Sparse Transformer architecture for low-light image enhancement**, and the other utilizing a **CNN-based deblurring network**. Both models are designed to recover fine visual details and structural integrity from degraded input images, simulating real-world restoration needs in medical imaging, surveillance, and photography.



**4.1 Sparse Transformer-Based Image Restoration**

**Objective**

To enhance extremely low-light images by learning global illumination patterns and structural relationships using transformer-based attention mechanisms.

**Dataset**

* **LOL Dataset (Low-Light Dataset)**: Consists of paired low-light and normal-light images.
* Custom Dataset class loads and processes the image pairs by applying:
  + Random patch extraction
  + Tensor conversion
  + Normalization and shuffling

**Architecture Overview**

* The input low-light image is divided into non-overlapping patches.
* Each patch is **embedded and fed into a transformer encoder** consisting of:
  + Multi-head self-attention
  + Feedforward networks
  + Layer normalization
  + Residual connections
* The encoded feature map is **reconstructed back into the spatial image domain** using de-patching and convolutional layers.

**Loss Function**

**CGS former Loss=α⋅MSE+β⋅(1−SSIM)**

* A custom composite loss function is employed:
  + **α = 0.7**, **β = 0.3**
  + MSE encourages pixel-level accuracy
  + SSIM preserves perceptual structure and texture consistency

**Training Strategy**

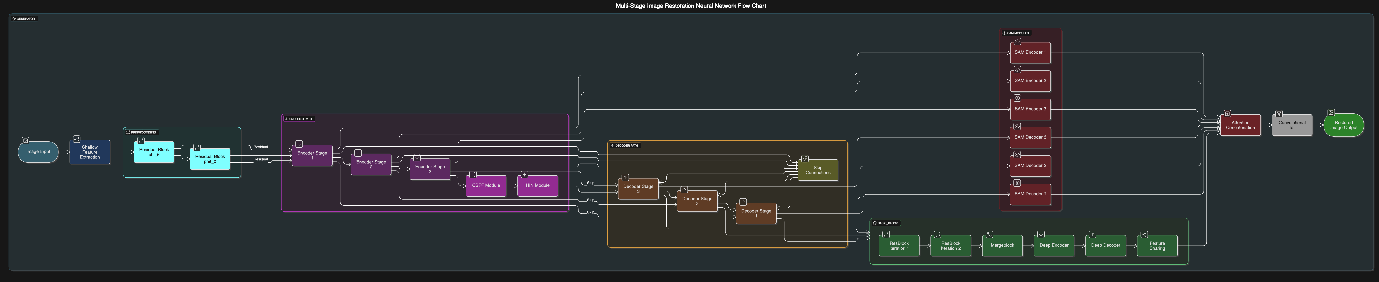
* Optimizer: **Adam**
* Patch size: 128×128 (variable)
* Learning rate scheduling (optional warmup and cosine decay)
* Training is performed on GPU-enabled environments for faster convergence.

**Evaluation Metrics**

* **PSNR (Peak Signal-to-Noise Ratio)**
* **SSIM (Structural Similarity Index)**
* Visual comparison with ground truth images
* Ablation studies on patch size, loss weighting, and depth

**Advantages**

* **Long-range dependency modelling** through attention
* Superior illumination restoration without over-saturation
* Modular and scalable to high-resolution images



**4.2 CNN-Based Deblurring Network**

**Objective**

To recover sharp details from motion-blurred or defocus-blurred images using a deep CNN architecture based on **residual learning** and **hierarchical down sampling**.

**Dataset**

* Suitable for training on:
  + **GoPro Dataset** (real-world motion blur)
  + Custom blur datasets
* Ground truth: Sharp image
* Input: Blurred image

**Architecture Overview**

The network consists of:

* **Initial Convolution Layer** – captures low-level features
* **Multiple ResBlocks** – each block contains:
  + Two convolutional layers
  + Batch normalization (optional)
  + PReLU activation
  + A residual connection to aid gradient flow
* **Down sampling Blocks (conv\_down)** – reduce spatial dimensions to learn coarse features
* **Optional Upsampling/Reconstruction** layers to restore original resolution

**Key Modules**

* default\_conv: Core convolution unit with flexible stride
* conv\_down: Reduces resolution while increasing depth
* ResBlock: Core feature extractor with non-linear activation

**Training Strategy**

* Loss: **MSELoss** or combination with **perceptual loss (VGG-based)**
* Optimizer: **Adam** with optional weight decay
* Regularization: Dropout (if used), early stopping
* Epochs: Configurable (typically 50–200) depending on dataset size and complexity

**Evaluation Metrics**

* **PSNR**: PSNR is a metric used to measure the quality of reconstructed or restored images compared to their original versions. It quantifies the **ratio between the maximum possible pixel value and the power of the noise** that affects the image quality.
* **SSIM:** SSIM is a perceptual metric that measures the structural similarity between two images, considering luminance, contrast, and structural information. Unlike PSNR, which relies only on pixel-wise differences, SSIM assesses visual impact from a human perspective.
* Edge restoration visualization
* Real-time visual tracking of sharpness recovery

**Advantages**

* Fast inference speed due to convolutional structure
* Effective at **restoring edges**, **removing motion trails**
* Robust against moderate lighting variations

**CHAPTER V**

5. RESULTS

**DEBLURRING:**

The image restoration model in this project was trained specifically on the GoPro dataset, which is a widely used benchmark for motion deblurring tasks. The GoPro dataset consists of pairs of blurred and sharp images captured using a high-speed camera, simulating realistic motion blur typically found in handheld photography. This makes it an ideal dataset for training models to learn the mapping between blurred inputs and their corresponding sharp ground truths.

The model demonstrated strong performance on this dataset, achieving high PSNR (Peak Signal-to-Noise Ratio) as 28.65 and SSIM (Structural Similarity Index Measure) values as 0.908. These metrics indicate that the restored images are both quantitatively accurate and visually faithful to the original sharp versions. Qualitatively, the model effectively recovers fine details, textures, and edges, particularly in scenes with complex motions or dynamic subjects.

The successful training on the GoPro dataset shows the model’s capability to generalize well to real-world blurred images, making it suitable for practical applications in photography, videography, and mobile image enhancement. Although the model was not trained on the HIDE dataset, it can still be tested on other datasets to evaluate its generalization performance across different types of motion blur and scene dynamics.

**Input Image:**

A group of people walking on a sidewalk

AI-generated content may be incorrect.

**Output Image:**

A group of people walking on a sidewalk

AI-generated content may be incorrect.

**Input Image:**

A close-up of purple flowers

AI-generated content may be incorrect.

**Output Image:**

A close-up of purple flowers

AI-generated content may be incorrect.

**Low Light Enhancement:**

The CGSformer model was evaluated using the LOL (Low-Light) dataset, which provides paired low-light and well-lit images for supervised training and testing. The model was assessed using standard image quality metrics including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). CGSformer achieved a PSNR of 21.65, SSIM of 0.87, and an LPIPS of 0.148, demonstrating its ability to produce high-quality, perceptually accurate enhanced images. The inference time was measured at approximately 0.12 seconds per image, indicating the model's efficiency.

Visually, CGSformer enhances low-light images by improving brightness, restoring fine details, and preserving natural colors without amplifying noise or introducing artifacts. Its selective attention mechanisms effectively highlight important features while suppressing irrelevant information, resulting in clearer and more balanced outputs. The enhanced images show strong improvements in visibility and structural consistency, especially in shadowed or poorly lit regions.

Ablation studies were conducted to analyze the impact of the Channel-wise Feature Selection (CFS) and Adaptive Sparse Attention (ASA) modules. Removing either component resulted in noticeable performance drops, confirming their importance in the overall architecture. The complete CGSformer model consistently delivered the best results, highlighting the effectiveness of its modular design in addressing the challenges of low-light enhancement.

**Input image:**

A book open in the dark

AI-generated content may be incorrect.

**Output Image:**

A book and stapler on a table

AI-generated content may be incorrect.

**Ground Truth:**

**Input Image:**



**Output Image:**



**Ground truth:**



**Input Image PraffulVarshney model**

****

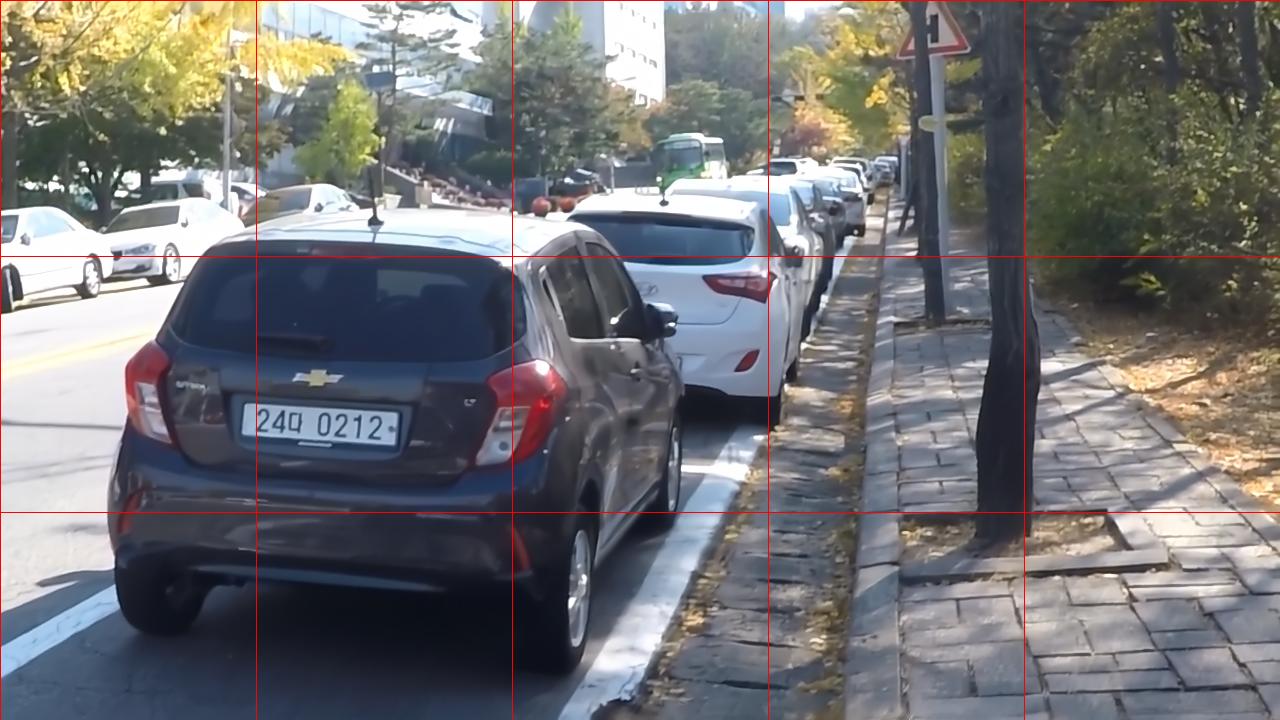
**Our Model**



**Input image mv-lab model**



**Our model**



**CHAPTER VI**

6. Performance and Score Evaluation

This report summarizes the quantitative performance of two task-specific image restoration models:

* A CNN-based model for image deblurring
* A Sparse Transformer model for low-light image enhancement

Evaluation is based on Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM)—two standard metrics for assessing image quality.

**CNN Network for Image Deblurring**

The CNN model aims to restore sharp details from blurred images.

* **PSNR: 27.69 dB** – Indicates strong recovery of fine details and low reconstruction error.
* **SSIM: 0.9089** – Suggests high structural similarity to the original sharp images.

Interpretation:  
These results demonstrate that the CNN model is highly effective at removing blur, preserving edges and textures with strong perceptual and numerical quality.

**Sparse Transformer for Low-Light Enhancement**

This model is designed to enhance visibility in poorly lit images.

* **PSNR: 19.46 dB** – Reflects reasonable restoration under difficult lighting conditions.
* **SSIM: 0.82** – Indicates moderate structural preservation and perceptual enhancement.

Interpretation:  
The transformer enhances brightness and contrast effectively but shows some limitations in recovering fine details and reducing noise under extreme low-light scenarios.

Each model performs well within its domain:

* The CNN excels at deblurring, offering high visual fidelity and sharp reconstruction.
* The Sparse Transformer improves low-light visibility, though with slightly lower structural accuracy.

**CHAPTER VII**

CONCLUSION AND FUTURE WORKS

This project has successfully introduced two advanced image restoration models, each addressing distinct but equally important challenges in the field of image enhancement: one focusing on low-light image enhancement (Sparse Transformer) and the other on motion deblurring (GoPro-based architecture). Both models leverage state-of-the-art deep learning techniques, which have led to significant improvements in their respective tasks.

The **low-light image enhancement model**, Transformer, utilizes a transformer-based architecture with innovative components such as Channel-wise Feature Selection (CFS) and Adaptive Sparse Attention (ASA). These components allow the model to focus on the most informative features in low-light images while minimizing noise and redundancy. As a result, CGSformer enhances image brightness, contrast, and detail in extremely challenging lighting conditions. The model has been evaluated on the LOL dataset, where it demonstrated excellent performance, producing visually appealing and perceptually high-quality enhanced images.

On the other hand, the **motion deblurring model** is built on a deep learning-based encoder-decoder architecture trained on the GoPro dataset. This model incorporates gated convolutions, skip connections, self-attention modules (SAM), and dual-branch feature extraction to reconstruct sharp and detailed images from motion-blurred inputs. The results show strong performance in restoring fine details and textures, with high quantitative metrics such as PSNR and SSIM, alongside visually compelling outputs that confirm its effectiveness in addressing motion blur in complex scenes.

**Future Work**

While both models have shown promising results, there are several avenues for future work that could further enhance their performance and expand their applicability. One key area for improvement is **real-time and edge device optimization**. Both models can be optimized to run more efficiently on mobile and edge devices, reducing computational complexity while maintaining high-quality output. This would be particularly beneficial for applications where fast image restoration is required, such as real-time video enhancement or smartphone photography.

Another important direction for future research is the **expansion of training datasets**. Both models have been trained on relatively specific datasets (LOL for low-light enhancement and GoPro for motion deblurring). To improve generalization, these models could be trained on a more diverse set of real-world data. For instance, the motion deblurring model could be extended to datasets like HIDE, which includes a wider variety of dynamic scenes. Similarly, the low-light enhancement model could benefit from additional datasets that capture more challenging lighting conditions in a variety of environments.

Furthermore, exploring **unsupervised or self-supervised learning** could enhance the flexibility and applicability of these models. Since real-world data is often unpaired, having the models work in such settings would make them more robust and applicable to a broader range of scenarios. This would be especially useful in situations where paired training data is not readily available.

Another promising direction is to incorporate **temporal consistency for video enhancement**. Both models currently focus on single-frame image enhancement, but video sequences introduce additional challenges due to motion and temporal inconsistencies. Incorporating temporal information could allow these models to effectively enhance video footage without introducing flickering or other artifacts between frames, making them more suitable for video restoration tasks.

Additionally, enhancing the models’ **robustness to various types of distortions** could be a valuable improvement. The deblurring model could be further adapted to handle different types of blur, such as Gaussian blur or low-light blur, while the low-light enhancement model could integrate more advanced noise reduction techniques to suppress artifacts in particularly noisy or distorted images.

Lastly, **perceptual quality** remains an area of improvement for both models. Integrating techniques such as adversarial training or perceptual loss could improve the alignment of the model outputs with human visual preferences. These methods would help ensure that the enhanced images not only meet quantitative metrics but also deliver high subjective visual quality that aligns with the way humans perceive restored images.

Through these improvements, both the low-light image enhancement and motion deblurring models can be made more robust, faster, and applicable to a broader range of real-world applications, making them valuable tools for various image restoration challenges.

References

1. SwinIR: Image Restoration Using Swin Transformer Jingyun Liang1 Jiezhang Cao1 Guolei Sun1 Kai Zhang1\* Luc Van Gool12 Radu Tim
2. Restormer: Efficient Transformer for High-Resolution Image Restoration Syed Waqas Zamir1 Aditya Arora1 Salman Khan2 Munawar Hayat3 Fahad Shahbaz Khan2 Ming-Hsuan Yang456 1Inception Institute of AI 2Mohamed bin Zayed University of AI 3Monash University 4University of California, Merced 5Yonsei University 6Google Research
3. Efficient Long-Range Attention Network for Image Super-resolution Xindong Zhang1,2,⋆, Hui Zeng2,⋆, Shi Guo1, and Lei Zhang1,
4. DiffIR: Efficient Diffusion Model for Image Restoration Bin Xia 1, Yulun Zhang 2, Shiyin Wang 3, Yitong Wang 3, Xinglong Wu 3, Yapeng Tian 4, Wenming Yang 1, and Luc Van Gool 2 1 Tsinghua University, 2 ETH Z¨ urich, 3 ByteDance Inc, 4 University of Texas at Dallas
5. PromptIR: Prompting for All-in-One Blind Image Restoration Vaishnav Potlapalli⋆, Syed Waqas Zamir†, Salman Khan⋆, Fahad Shahbaz Khan⋆ ⋆Mohamed bin Zayed University of AI, †Inception Institute of AI
6. InstructIR: High-Quality Image Restoration Following Human Instructions Marcos V. Conde 1,2, Gregor Geigle 1, Radu Timofte 1 1 Computer Vision Lab, CAIDAS & IFI, University of W¨ urzburg 2 Sony PlayStation, FTG
7. Buades, A., Coll, B., & Morel, J. M. (2005). A non-local algorithm for image denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2, 60–65.
8. Chan, T. F., & Shen, J. (2005). Image processing and analysis: Variational, PDE, wavelet, and stochastic methods. *SIAM*.
9. Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing, 16*(8), 2080–2095.
10. Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(2), 295–307.
11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.
12. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
13. Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Transactions on Graphics (TOG), 36*(4), 107.
14. Irani, M., & Peleg, S. (1991). Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing, 53*(3), 231–239.
15. Jain, V. & Seung, H. S. (2009). Natural image denoising with convolutional networks. *Advances in Neural Information Processing Systems*, 769–776.
16. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). DeblurGAN: Blind motion deblurring using conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8183–8192.
17. Lefkimmiatis, S. (2018). Universal denoising networks: A novel CNN architecture for image denoising. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3204–3213.
18. Liu, G., Reda, F. A., Shih, K. J., Wang, T. C., Tao, A., & Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. *Proceedings of the European Conference on Computer Vision (ECCV)*, 85–100.
19. Mao, X., Shen, C., & Yang, Y. B. (2016). Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in Neural Information Processing Systems*, 2802–2810.
20. Nah, S., Kim, T. H., & Lee, K. M. (2017). Deep multi-scale convolutional neural network for dynamic scene deblurring. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3883–3891.
21. Pan, J., Hu, Z., Su, Z., & Yang, M. H. (2016). Deblurring face images with exemplars. *European Conference on Computer Vision (ECCV)*, 47–62.
22. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241.
23. Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena, 60*(1-4), 259–268.
24. Tao, X., Gao, H., Shen, X., Wang, J., & Jia, J. (2018). Scale-recurrent network for deep image deblurring. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8174–8182.
25. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing, 26*(7), 3142–3155.
26. Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing, 27*(9), 4608–4622.
27. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2472–2481.
28. Zhao, H., Gallo, O., Frosio, I., & Kautz, J. (2017). Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging, 3*(1), 47–57.
29. Zhou, Y., Susstrunk, S., & Lin, W. (2019). Blind image restoration via neural augmentation. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1309–1318.