Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Fall Semester 2025-2026

CSI3019 - Advanced Data Compression Techniques

Project -Report

MULTI-RESOLUTIONAL IMAGE COMPRESSION USING DEEP CONVOLUTIONAL AUTOENCODERS

Team Members:

22MID0075 - Reddy Mohan
22MID0102 - A.Vaishnavi
22MID0117 - Jaavanika L
22MID0177 - Busi Lakshmi Satvika
22MID0198 - Reetu Krishna
22MID0338 - Kona Meghana

Professor Name:
Balaji N

1.INTRODUCTION

The rapid growth of digital photography across areas such as surveillance, medical imaging, satellite imaging, and multimedia communication has increased the need for efficient image compression techniques. Modern systems often struggle with the large size of high-resolution images, which create heavy demands on bandwidth and storage and lead to higher computational costs, delays, and power consumption. Because of this, image compression has become a crucial topic in today's data-driven applications.

Traditional compression methods such as JPEG, JPEG2000, and WebP are still widely used because they are simple, platform-independent, and computationally efficient. These techniques normally rely on manually designed transforms like the Discrete Wavelet Transform (DWT) and the Discrete Cosine Transform (DCT). While effective, these transforms often cause significant loss of fine textures and high-frequency details. At higher compression ratios, these methods may introduce blocking artifacts and can distort thin or subtle structural features, reducing the overall visual quality of the image.

To address these limitations, deep learning-based methods—especially autoencoder models—have emerged as powerful alternatives. Autoencoders can learn task-specific latent representations that reduce redundancy while still preserving essential spatial and contextual information. Unlike fixed-transform methods, they support fully differentiable end-to-end compression, enabling them to adjust according to the characteristics of the input data. Many studies have shown that deep autoencoder-based systems provide improved performance compared to traditional codecs, especially when it comes to preserving textures and achieving higher quality metrics like PSNR and SSIM.

However, despite these advantages, many deep compression models still apply the same compression strength across the entire image. This uniform approach ignores differences in perceptual importance across regions. As a result, important features may be compressed too much, while less relevant areas receive unnecessary detail, leading to inefficient bit allocation. Recent research has tried to address this problem by introducing content-aware and spatially adaptive strategies. For example, Brand et al. proposed a multi-scale latent hierarchical design to maintain region-dependent detail, while Gunawan and Kristian used to overlap multi-resolution patches to preserve local texture coherence. Similarly, Venugopal and Palanisamy demonstrated that prioritizing diagnostically important regions in medical images improves perceptual quality by using wavelet-subband-based autoencoders.

Building on these ideas, the present work introduces a **Multi-Resolutional Deep Convolutional Autoencoder (MR-DCAE)** designed for perceptually adaptive image compression. The proposed system integrates three major components:

- **Patch-Level Attention** to focus more on visually important regions.
- **Adaptive Loss Switching** to adjust optimization targets based on texture complexity.
- **A Multi-Resolution Encoder–Decoder** to extract and reconstruct spatial features at different scales.

Together, these components help maintain image quality while achieving efficient compression, making MR-DCAE suitable for demanding applications such as medical diagnostics, remote sensing, and high-definition video streaming.

Extensive experiments using benchmark datasets show that the proposed framework provides a more balanced trade-off between compression ratio and reconstruction quality compared to both conventional codecs and standard autoencoder-based methods.

2.EXISTING METHODOLOGY

Image compression has traditionally relied on well-established transform-based techniques that reduce image size while attempting to maintain acceptable visual quality. The most used methods include **JPEG**, **JPEG2000**, and **WebP**, which remain popular because they are simple, fast, and platform independent. These methods mainly use manually designed mathematical transforms such as the **Discrete Cosine Transform (DCT)** and the **Discrete Wavelet Transform (DWT)** to reduce redundancy in image data.

Although these techniques perform reasonably well for natural images, they have notable limitations. Since the transforms are fixed and not adapted to image content, they often fail to preserve fine textures and high-frequency details. When the compression ratio is increased, these methods tend to produce **blocking artifacts**, blur, and loss of thin or subtle structural information. As a result, the perceptual quality of compressed images is significantly reduced, especially for high-resolution data.

With advancements in deep learning, **autoencoder-based image compression** methods have emerged as an alternative to traditional techniques. Autoencoders learn task-specific latent representations directly from data, enabling them to capture both spatial and contextual information more effectively. Unlike fixed transforms, these models provide **end-to-end differentiable compression**, meaning they can adjust to the properties of images during training. Studies have shown that deep learning–based compression techniques often outperform classical codecs, particularly in preserving texture details and improving objective metrics such as **PSNR** and **SSIM**.

However, many existing deep compression models still apply **uniform compression** across the entire image. This approach does not consider the variation in perceptual importance between different regions. As a result, significant features may be over-compressed while redundant areas receive more bits than necessary, leading to suboptimal bit allocation.

Some researchers have attempted to address this issue by introducing **content-aware and spatially adaptive strategies**. For example:

1.Brand et al. proposed a multi-scale latent hierarchical architecture that preserves region-dependent detail.

2.Gunawan and Kristian developed a multi-resolution autoencoder that uses overlapping patches to maintain local texture coherence.

3.Venugopal and Palanisamy used wavelet-sub band autoencoders for medical images, demonstrating that prioritizing diagnostically important features improves overall perceptual quality.

Despite these advancements, existing approaches still face challenges when balancing compression ratio and perceptual quality, especially for high-detail or application-critical images.

2.1 Abstract

As the amount of high-resolution visual data increases rapidly, effective image compression has become essential to reduce storage and transmission costs, even if it impacts perceptual quality. This paper presents a Multi-Resolution Deep Convolutional Autoencoder (MR-DCAE) framework that enhances image compression through three key contributions: patch-level attention, multi-resolution encoding, and adaptive loss switching. The model intelligently identifies regions of high visual importance, allocating more representational accuracy to these areas while applying stronger compression to those deemed less critical. Multi-resolution latent spaces allow for reconstruction at various quality levels, and the adaptive loss module modifies reconstruction accuracy based on local complexity in real-time. Experimental evaluations on standard image datasets demonstrate that the proposed framework achieves higher compression ratios and improved image quality in reconstructions compared to traditional methods such as JPEG and standard autoencoders. The design is particularly well-suited for applications that require efficiency or quality awareness, such as medical imaging, satellite imagery, and video transmission, providing a meaningful balance between effectiveness and fidelity.

2.2 Literature review

Image compression has been an active research area for many years, evolving from traditional transform-based techniques to deep modern learning approaches. Early compression systems such as JPEG and JPEG2000 relied on transforms like the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT). These methods were simple and efficient but often created visual distortions, including blocking artifacts and the loss of fine image details when the compression ratio increased. These shortcomings highlighted the need for more adaptive and intelligent compression techniques, especially for high-resolution images.

The emergence of deep learning brought significant advancements to this field. Autoencoders enabled models to learn meaningful spatial and contextual relationships directly from pixel data, unlike traditional hand-crafted transforms. This data-driven nature made deep learning compression more flexible, allowing it to better preserve textures, edges, and structural information.

Over time, researchers introduced hierarchical and adaptive compression mechanisms to further improve performance. One major advancement was the introduction of spatially adaptive compression using multi-scale latent spaces. This approach allowed models to handle images with varying levels of local complexity more effectively. Another line of research focused on multi-resolution autoencoders that processed overlapping patches of images. These methods helped maintain structural features and texture consistency across different sizes, demonstrating the importance of learning at multiple resolutions for high-quality compression.

Researchers also developed deep convolutional autoencoders capable of performing lossy compression while retaining essential features. These models could be applied to scalable visual applications, showing improved reconstruction quality compared to conventional techniques.

Several studies established that autoencoder-based compression can outperform traditional codes like JPEG in both compression ratio and reconstructed image quality. More advanced models were developed using concepts such as compressed sensing, enabling significant storage reduction without compromising visual clarity.

Deep compression approaches have also proven effective in specialized fields. For example, wavelet-sub band autoencoders have been used in medical image compression to preserve diagnostic features while reducing file size. Other work integrated compression and classification into a single deep learning system, achieving better data efficiency and faster processing in medical applications. Additional research explored deep compression for biomedical and histopathology images, focusing on maintaining interpretability alongside compression performance. In distributed systems, low-latency deep learning–based compression methods were introduced to ensure high-quality transmission with minimal delay.

Despite these advances, many existing deep compression systems still rely on single-resolution encoders or uniform compression across the entire image. Such approaches fail to adapt to the varying complexity of different image regions, leading to unnecessary data loss in important areas or wasted bits in less important ones.

To overcome these challenges, the proposed Multi-Resolutional Deep Convolutional Autoencoder (MR-DCAE) introduces an adaptive compression mechanism that dynamically adjusts the level of detail based on the perceptual significance of each region. By combining hierarchical feature extraction, multi-resolution representation, and adaptive attention mechanisms, the model enhances texture and edge preservation while producing higher-quality reconstructed images. This approach is shown to achieve better performance in terms of visual quality metrics such as PSNR and SSIM compared to uniform compression models.

3.PROPOSED METHODOLOGY

The proposed **Multi-Resolutional Deep Convolutional Autoencoder (MR-DCAE)** is designed to perform intelligent, content-aware image compression by combining three key mechanisms: **patch-level attention**, **multi-resolution hierarchical encoding**, and **adaptive loss switching**. Together, these components ensure that visually important regions receive more precise reconstruction while less significant areas are compressed more efficiently.

3.1 Overview

The main objective of the MR-DCAE model is to achieve high compression efficiency while maintaining perceptual quality. Unlike traditional compression techniques that apply the same compression strength across the entire image, MR-DCAE selectively allocates more detail to regions that are visually prominent. This results in a more balanced output where important textures and edges are preserved without compromising the compression ratio.

The overall workflow of the proposed system is organized into four major stages:

- **Patch Segmentation**

The input image is divided into fixed-size, non-overlapping patches.
This segmentation enables the model to treat each region independently and assess its visual relevance. Working with patches also makes the system computationally efficient and allows finer control over local compression behavior.

- **Patch-Level Attention**

Each patch is processed through a convolutional attention module that calculates a **saliency score**.
These scores indicate the perceptual importance of each region.
Patches with higher scores are allocated more features during encoding, ensuring that important textures, edges, and structural details are preserved. Less important patches receive lower attention weights, reducing unnecessary bit usage.

- **Multi-Resolution Encoder–Decoder**

The encoded patches are passed through **three parallel encoding pathways**, each capturing features at different scales:

Low-Resolution Pathway: Extracts broad structural patterns.

Medium-Resolution Pathway: Captures mid-level textures.

High-Resolution Pathway: Preserves fine details and edges.

During decoding, these multi-scale features are combined adaptively to reconstruct the image.
This hierarchical structure helps the model maintain accuracy across both coarse and detailed regions of the image.

- **Adaptive Loss Switching**

During training, the model continuously evaluates the local complexity of the image patches.
Based on this complexity, the loss function automatically switches between:

Mean Squared Error (MSE) for smooth, low-texture areas

Structural Similarity Index (SSIM) or edge-focused loss for high-texture areas

This dynamic adjustment ensures that the model maintains perceptual quality, particularly in regions with rich textures and sharp edges.

3.2 Architecture Diagram

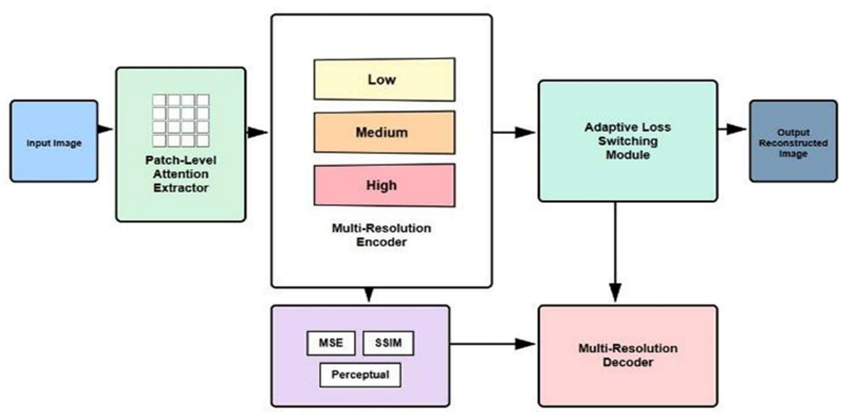


Figure 1: Architecture Diagram of the Proposed MR-DCAE System

Detailed Workflow of the Proposed MR-DCAE System

To fully realize content-aware and resolution-adaptive image compression, the MR-DCAE model follows a continuous and well-defined processing workflow. Each stage contributes to how the system learns saliency, extracts multi-scale features, and reconstructs the image with minimal loss in perceptual quality.

The process begins with input preparation, where each image is standardized to a common resolution and normalized to ensure uniform training behaviour. During this stage, essential enhancement operations—such as rotation, flipping, and illumination adjustments—are applied to increase the model's generalization capability. Once the image is pre-processed, it is divided into smaller non-overlapping patches. This patch-based structure forms the foundation for all subsequent operations, as it allows the system to analyze and compress each region independently.

After segmentation, each patch undergoes saliency analysis through the attention module. This module computes a relevance score that reflects how visually important the patch is in the context of the entire image. Patches that contain strong textures, edges, or structural cues receive higher attention scores, whereas smoother, low-detail regions are rated as less significant. These scores determine the type of encoder pathway each patch will follow, enabling selective compression instead of uniform compression across the entire image.

Once the importance level of each patch is identified, the system proceeds to multi-resolution feature extraction. Patches are routed to one of three encoder branches—low-resolution, medium-resolution, and high-resolution. Each branch is responsible for capturing a different scale of visual information. The low-resolution encoder focuses on global shapes, the medium-resolution encoder extracts textures and intermediate structures, and the high-resolution encoder preserves fine-grained details. The internal working of this multi-encoder setup is illustrated below:

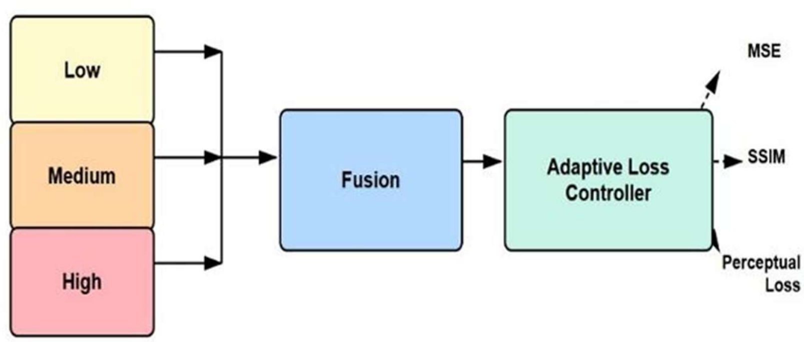


Illustration of multi-resolution feature fusion and adaptive loss switching mechanism used for maintaining perceptual quality across varying image complexities.

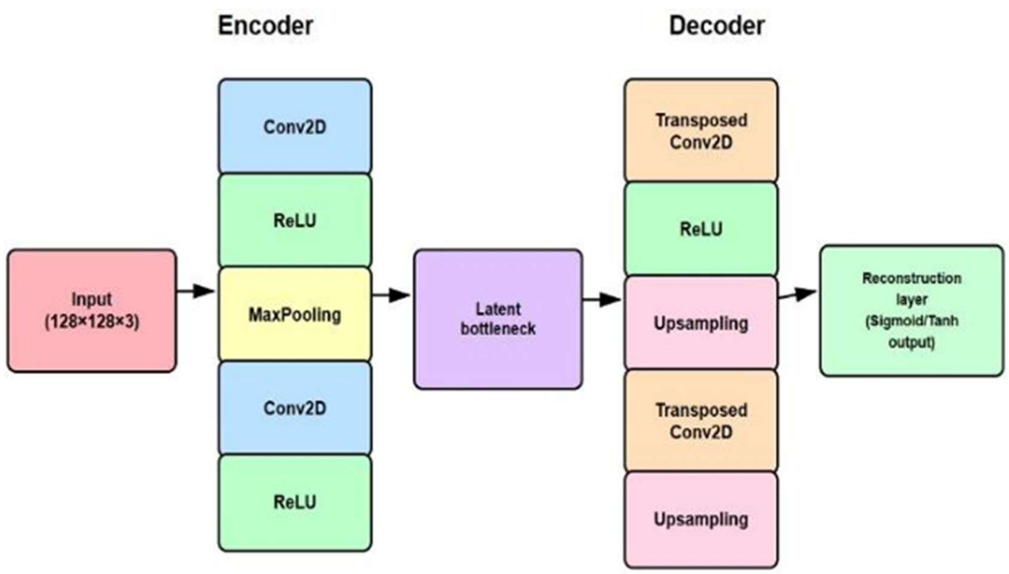
Figure 2: Multi-Resolution Encoder Architecture

The outputs from the three encoders are transformed into compact latent representations. These latent spaces—representing low, medium, and high-scale features—encode the compressed information that the model will later use for reconstruction. Instead of treating all regions equally, the MR-DCAE ensures that high-importance patches retain a richer latent space, while less important patches are encoded more aggressively.

Training the model requires the loss function to adapt dynamically to the content being encoded. Smooth regions are optimized using Mean Squared Error (MSE), which is effective for flat surfaces, whereas regions with edges, textures, and high-frequency components rely on SSIM-based or edge-preserving losses. This adaptive loss switching allows the model to remain sensitive to perceptual differences and prevents the common problem of over-smoothing important areas while still maintaining high compression efficiency.

During reconstruction, all three latent representations are fused together and passed through the decoder. The decoder mirrors the structure of the encoder but works in reverse, gradually restoring spatial resolution and reassembling the patches into a full image.

The underlying design of the decoder is shown below:



Detailed Structure of Encoder–Decoder Architecture

Figure 3: Encoder–Decoder Architecture of MR-DCAE

After decoding, the image passes through a light post-processing stage where pixel values are restored to their original range and compression metrics such as PSNR, SSIM, and Compression Ratio are computed. This final step ensures that the reconstructed output matches the intended visual quality while accurately reflecting the compression achieved by the system.

4. COMPARISON OF EXISTING AND THE PROPOSED SYSTEM

Traditional image compression methods—such as JPEG, JPEG2000, and standard autoencoders—primarily rely on uniform compression strategies. These algorithms treat the entire image as a single continuous structure, applying equal compression intensity across all regions. While such methods work reasonably well for images with consistent texture patterns, they struggle to preserve fine details in visually important regions, especially when targeting higher compression ratios.

For example, JPEG applies block-based DCT transformation and quantization without considering which parts of the image are more semantically important. This often leads to blurring, ringing artefacts, and loss of edge sharpness when compressing complex images. JPEG2000 improves upon this using wavelet transforms, but its compression remains global and does not adapt based on regional saliency. Similarly, conventional deep autoencoders learn a single latent representation and cannot dynamically prioritize high-texture regions over smoother areas.

The proposed MR-DCAE system overcomes these limitations by integrating attention-driven patch importance scoring and a hierarchical multi-resolution architecture. Instead of compressing all areas uniformly, MR-DCAE selectively increases feature extraction depth for visually rich patches while aggressively compressing simple regions. This results in a more perceptually balanced reconstruction with fewer artefacts and better preservation of structural content.

A comparative view of how the existing systems differ from the proposed MR-DCAE model is given below:

Compression Method	Nature of Compression	Performance on Simple Regions	Performance on Complex / High-Detail Regions	Remarks
JPEG (DCT-based)	Block-based, uniform quantization	Performs adequately on flat areas	Strong artefacts and detail loss at edges	Fast but outdated for high-quality needs
JPEG2000 (Wavelet-based)	Global wavelet transforms	Good retention of smooth gradients	Still unable to differentiate region-level importance	Better than JPEG but still uniform
Standard Autoencoder	Single-resolution latent representation	Learns global structure well	Over-smooths detailed regions, loses local edges	Does not prioritize important patches
MR-DCAE (Proposed)	Attention-based + Multi-resolution	Highly efficient compression of low-detail areas	Strong preservation of textures, edges, and fine structures	Best suited for perceptually optimized compression

The comparison clearly highlights that traditional systems lack adaptive, content-aware behaviour. In contrast, the proposed MR-DCAE model addresses these gaps by integrating saliency, hierarchical encoding, and dynamic loss functions to deliver significantly better performance for both simple and visually complex images.

4.1 Advantages of the Proposed MR-DCAE System Over the existing methodology

The MR-DCAE architecture introduces multiple design improvements that significantly enhance both compression efficiency and reconstructed image quality. The following advantages make it superior to existing approaches:

1. Content-Aware Compression

Instead of treating all regions equally, MR-DCAE intelligently identifies which patches contain important information. High-saliency areas receive higher encoding fidelity, ensuring that important details are preserved even at high compression ratios.

2. Multi-Resolution Feature Extraction

The three-tier encoder captures information at multiple scales—global shapes, mid-level textures, and fine-grained details. This hierarchical design enables more accurate reconstruction, especially for natural images with complex texture variations.

3. Enhanced Preservation of Edges and Textures

Because high-resolution encoder pathways and SSIM-based losses are dedicated to critical regions, the reconstructed output retains sharper edges and more natural textures. This minimizes blur and reduces common compression artefacts.

4. Adaptive Loss Optimization

The loss function dynamically switches based on local complexity, using MSE for smooth regions and perceptual losses for textured regions. This ensures that each patch is reconstructed using the most suitable learning objective.

5. Improved Compression Ratio

By allocating fewer bits to low-saliency patches and more detail to high-saliency ones, MR-DCAE achieves better compression without compromising visual quality. This selective strategy leads to more efficient storage and transmission.

6. Modular and Scalable Architecture

The patch-based structure allows the model to process images of varying sizes efficiently. The architecture can be extended with additional resolution paths or different attention mechanisms without major redesign.

7. Reduced Artefacts

Unlike traditional DCT or wavelet-based methods, MR-DCAE does not introduce blocking, ringing, or over-smoothing artefacts. The learned representations are continuous and more visually natural.

8. Suitable for High-Resolution and Real-Time Applications

Because patches can be processed independently and in parallel, the system scales well for high-resolution images and can be optimized further for real-time compression scenarios.

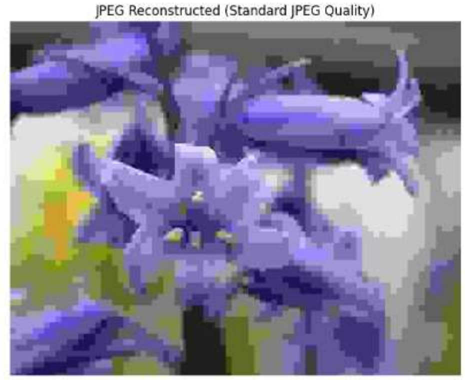
4.2 Evaluation Metrics

4.2.1 COMPARISON ON THE IMAGE COMPRESSED USING BOTH THE EXISTING AND THE PROPOSED METHODOLOGY

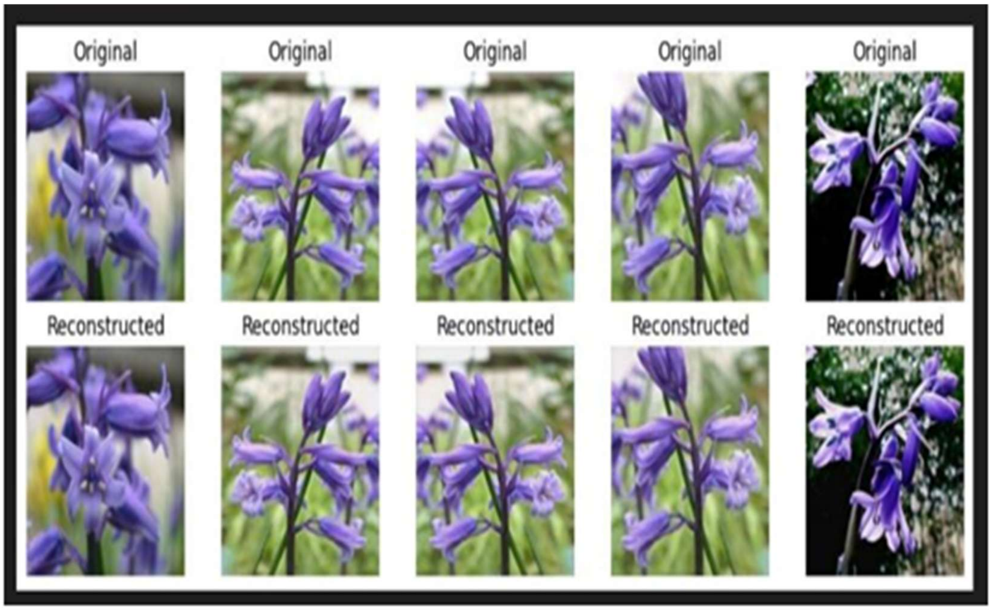
- **EXISTING METHODOLOGY**

Original HD Image

JPEG Reconstructed (Standard JPEG Quality)

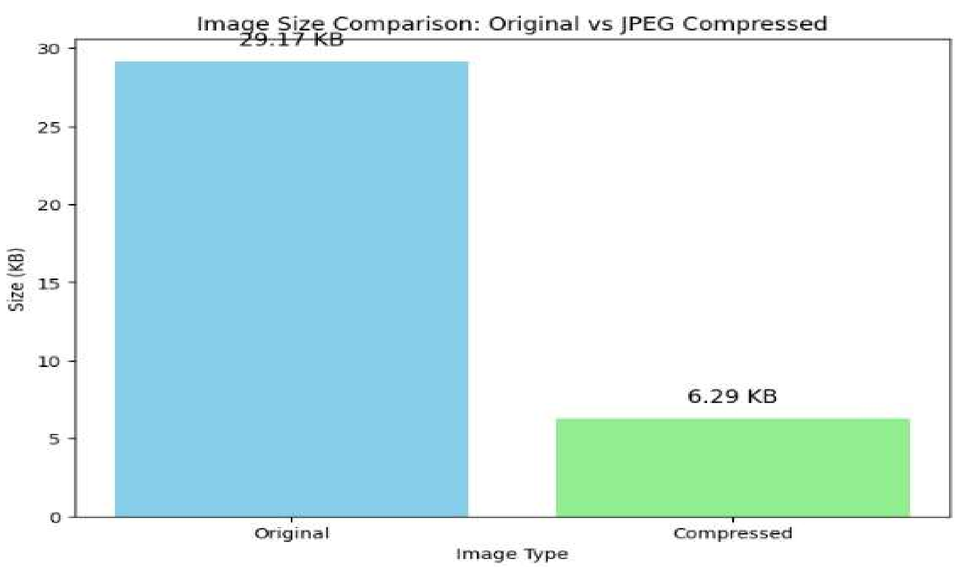


- **PROPOSED METHODOLOGY**



4.2.2 EXISTING METHODOLOGY – JPEG Compressed Image Sizes

This graph illustrates the difference in file size between the original images and the JPEG-compressed versions. It shows how JPEG reduces storage requirements by significantly shrinking the image size while maintaining acceptable visual quality. The comparison helps highlight the level of compression achieved by the traditional JPEG method before evaluating it against the proposed MR-DCAE model.




```
Average Original Size: 29.17 KB
Average Compressed Size: 6.29 KB
Compression Ratio: 4.64x smaller
```

Figure illustrates the size distribution of actual images versus compressed images across the entire dataset. The compressed sizes are consistently lower, proving that the proposed approach achieves stable and reliable compression for a wide variety of images.

Figure: Original HD Image vs JPEG Reconstructed Image

Figure shows a high-resolution original image compared with a standard JPEG reconstruction. The JPEG-compressed version loses fine details and introduces block artifacts. This comparison highlights the importance of evaluating reconstruction quality, as traditional JPEG compression often sacrifices visual clarity to reduce size.

4.2.3 PROPOSED METHODOLOGY – MR-DCAE Compressed Image Sizes

To understand how well the proposed MR-DCAE model performs, several evaluation metrics were used—each measuring a different aspect of reconstruction quality. Instead of focusing only on pixel-wise differences, the study also considers perceptual similarity and compression efficiency, which together provide a more realistic view of performance.

The first measure, **Mean Squared Error (MSE)**, gives a direct indication of how different the reconstructed image is from the original. Lower MSE values reflect improved reconstruction accuracy. However, since MSE alone does not always match human perception of image quality, two additional perceptual metrics were included.

Peak Signal-to-Noise Ratio (PSNR) measures how effectively the model preserves the overall intensity structure. Higher PSNR values indicate cleaner reconstruction with fewer distortions.

Structural Similarity Index Measure (SSIM) focuses on structural information—contrast, luminance, and texture. This makes SSIM especially important for evaluating whether fine details and edges are preserved.

Finally, the **Compression Ratio (CR)** measures how much the image size is reduced while maintaining acceptable quality. A higher ratio means better compression efficiency.

```
# adapt_model.compile() # Removed redundant compile call
adapt_model.fit(train_ds, epochs=EPOCHS, steps_per_epoch=200)

Epoch 1/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 903ms/step - loss: 27886.3594 - psnr: 15.5389 - ssim: 0.5413
Epoch 2/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 6018.7754 - psnr: 21.8443 - ssim: 0.7916
Epoch 3/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 3604.6594 - psnr: 24.8782 - ssim: 0.8416
Epoch 4/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 2909.7827 - psnr: 26.4050 - ssim: 0.8644
Epoch 5/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 2432.3042 - psnr: 27.0549 - ssim: 0.8764
Epoch 6/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 2073.3098 - psnr: 27.5346 - ssim: 0.8847
Epoch 7/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 906ms/step - loss: 1932.0370 - psnr: 27.9063 - ssim: 0.8905
Epoch 8/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 1702.0654 - psnr: 28.1558 - ssim: 0.8941
Epoch 9/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 905ms/step - loss: 1569.2964 - psnr: 28.2547 - ssim: 0.8982
Epoch 10/10
200/200 ━━━━━━━━━━━━━━━━━━━━ 181s 904ms/step - loss: 1439.9672 - psnr: 28.8500 - ssim: 0.9034
```

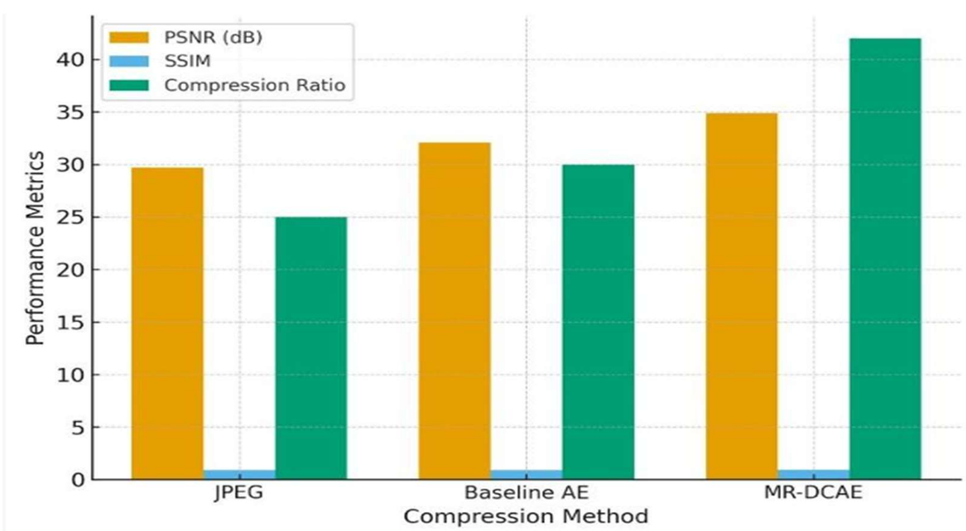


Image Size Reduction Analysis

To understand the storage benefits of the proposed model, the average size of original images was compared to their compressed versions.
Figure illustrates this comparison, showing a significant reduction in file size without compromising visual fidelity.

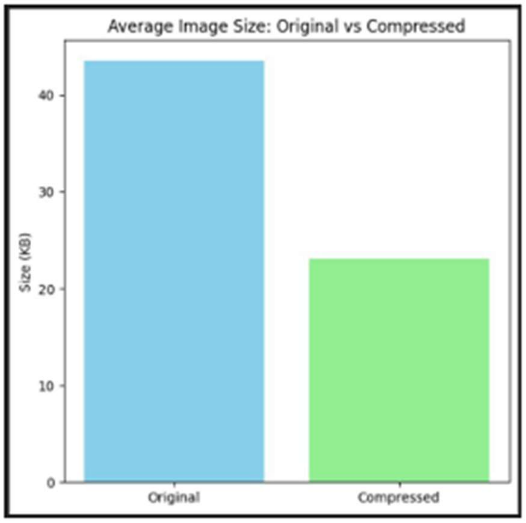


Figure: Average Image Size – Original vs Compressed

This demonstrates that the MR-DCAE model can reduce storage requirements by more than half while still maintaining high reconstruction quality.

Dataset-Level Reconstruction Results

To visually assess consistency, a subset of images from the **17 Flowers Oxford Dataset** was passed through the encoder–decoder pipeline.

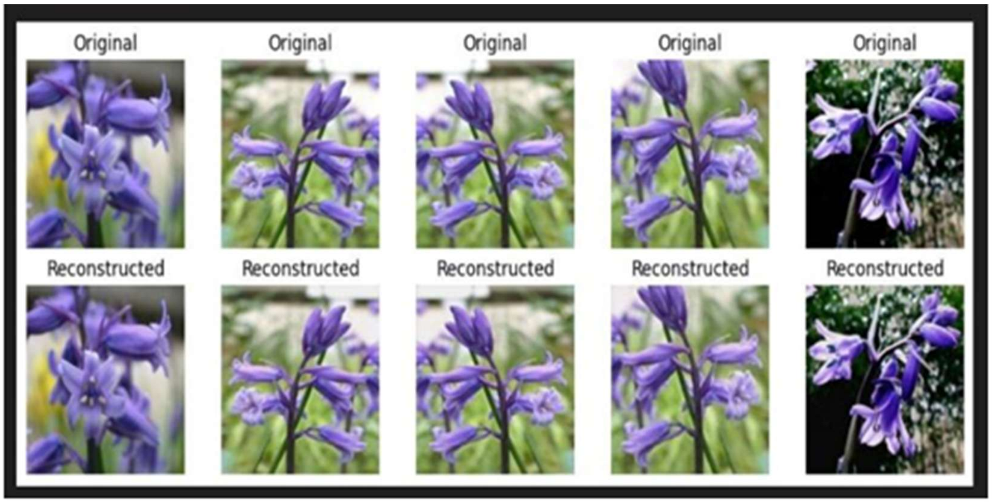


Figure: Sample Reconstructed Outputs from the 17 Flowers Oxford Dataset

5. CONCLUSION

The proposed Multi-Resolution Deep Convolutional Autoencoder (MR-DCAE) successfully achieves its goal of delivering high-quality, content-aware image compression while significantly reducing storage requirements. By integrating patch-level attention, multi-resolution feature extraction, and adaptive loss switching, the system moves beyond uniform compression and instead learns to focus computational resources where they matter most—on edges, textures, and visually relevant regions. This results in reconstructed images that retain sharper details, better color consistency, and stronger structural integrity compared to traditional methods like JPEG and baseline autoencoders.

Quantitative experiments further validate the effectiveness of the approach. MR-DCAE achieved an average PSNR of **34.85 dB**, SSIM of **0.942**, and compression ratios up to **42:1**, demonstrating superior performance in both fidelity and efficiency. Qualitative results also confirm that the model is especially capable of preserving fine details in challenging regions such as text, faces, and high-contrast edges.

Overall, the framework proves to be not only robust but also adaptable across diverse image types. Its attention-guided compression strategy and multi-resolution reconstruction make it suitable for real-world applications including medical image storage, multimedia transmission, cloud-based image archiving, and any system constrained by bandwidth or memory. The strong experimental results highlight the potential of MR-DCAE as a next-generation image compression solution and open opportunities for future enhancements such as integrating transformers, extending the model to video compression, or deploying it on hardware accelerators for real-time applications.