Compressed Image Super Resolution using Convolutional Neural Network

Ei Ei Tun
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Bangkok 10330, Thailand
ei.ei2@alumni.chula.ac.th

Aktanin Konkitkriengkrai
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Bangkok 10330, Thailand
boomatn@gmail.com

Watchara Ruangsang
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Bangkok 10330, Thailand
watchara.ru@student.chula.ac.th

Supavadee Aramvith
Multimedia Data Analytics and Processing Research Unit
Department of Electrical Engineering, Chulalongkorn University
Faculty of Engineering
Bangkok 10330, Thailand
Supavadee.A@chula.ac.th

***Abstract*—Image compression is a topic of significant interest as it reduces file sizes in stored data. In this paper, we propose a model that achieves multiple levels of compression, thereby minimizing the storage space required for images, which typically consume substantial amounts of data due to their size and resolution. We combine an image downscaling and upscaling model with an image compression model. By leveraging convolutional techniques to identify image features, we can effectively reduce the size of the image through downscaling and subsequently upscaling it. Additionally, we employ entropy image compression and arithmetic encoding to compress and reconstruct the image while preserving its lossless data. Through experimentation with the Kodak dataset, we observed that our proposed model achieved a compression rate of 96.92%, significantly reducing the data needed for file storage. Moreover, our reconstructed images attained a standardized measure with a signal-to-noise ratio of 33.10 dB and a structural similarity of 0.9219. Notably, the perceptual quality of the images, including intricate details, remained intact to the human eye.**

***Index Terms*—downscaling, compression, image reconstruction, the entropy model**

I. Introduction

Image compression plays a significant role nowadays, especially as technology advances [1]. When people desire clearer and higher-quality images, it requires more storage space for data storage. This leads to excessive storage consumption, whether on hard drives or software. If we can reduce the file size of images, it would lower the storage costs. This is why a growing interest is in high-quality image compression techniques that can reconstruct images. Therefore, this work was created to address these issues.

This work focuses on compressing high-quality images using advanced techniques discussed in the literature [2] and [3]. Literature [2] introduces a model for compressing and reconstructing images with excellent detail using non-lossy data compression methods. On the other hand, literature [3] explores techniques for reducing image size and restoring images through upscaling. Both models work together to overcome the limitations of using compression alone, which may result in insufficient detail. In this work, we combine the models from the literature [2] and [3] to achieve highly compressed images with preserved detail and the ability to reconstruct them.

The expected outcome of the work is to develop a model that can compress high-resolution images and restore them effectively. This will benefit practical applications, especially in engineering industries requiring detailed images and efficient storage. The work uses advanced compression techniques to reduce storage costs while maintaining excellent performance. Each model has different compression capabilities and excels in image restoration. The work's approach is straightforward and practical, focusing on simplicity and usability.

The anticipated outcomes of the research work are

- Develop a model that can compress and restore high-resolution images effectively.
- Test and compare different models for their compression capabilities.
- Evaluate how well each model compresses images and the storage space required.
- Test the selected models' ability to restore images.
- Create a practical and straightforward approach for image compression and restoration.

The rest of this paper is organized as follows. Section II and III describes the details of the model architecture and evaluation of the proposed model, respectively. In the end, Section IV discusses, and Section V concludes our paper, respectively.

II. Methodology and Analysis

This section describes the details of the model architecture and the analytical results in each model variation.

A. The details of the model architecture

In Fig.1, the integrated model's workflow is described as follows. The model takes a high-resolution (HR) image as

This research is funded by Thailand Science research and Innovation Fund Chulalongkorn University (CU_FRB65_ind (9)_157_21_23) and (IND66210019) and the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research, and Innovation [grant number B04G640053].

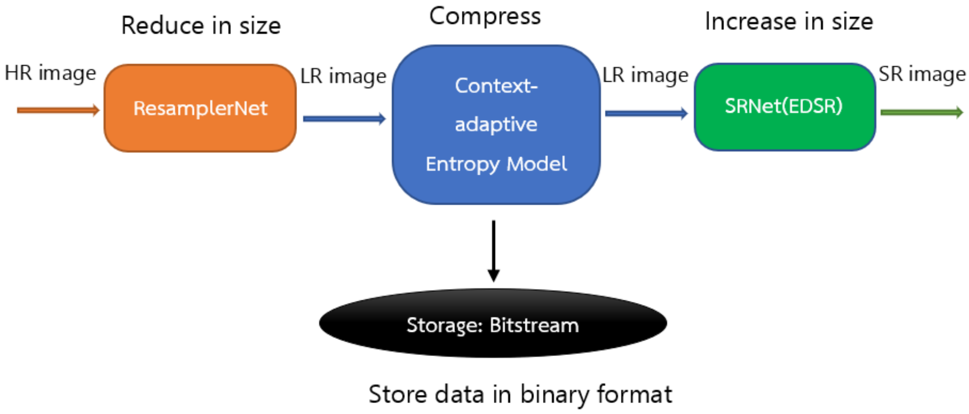

Fig. 1. Steps of the integrated model's workflow.

input. The image is resized using a Resampler Network, which predicts and places downsampled random samples on the corresponding pixels of the high-resolution image. This creates a low-resolution (LR) image. The low-resolution image is compressed using a hybrid context-adaptive entropy model assisted by hyperprior, which predicts and generates compressed bitstreams. The compressed data is stored in a binary format with the .cmp file extension. To reconstruct the image, the compressed data in the binary format is decoded to retrieve the low-resolution image. Finally, the low-resolution image is upscaled using an EDSR (Enhanced Deep Super-Resolution) process, which employs progressive development and sub-pixel convolution to increase the image size while preserving details. In summary, the combined model follows a step-by-step plan that integrates a Resampler Network for reducing image size, a hybrid context-adaptive entropy model for compressing data, and an EDSR process for enhancing image quality. This careful sequence improves image analysis while keeping important details intact.

B. Results of Each Model

This paper used two models, each focusing on different aspects of image processing. The first model is a size reduction/enlargement model, while the second is a compression model. The specific models used in this project are the High-Resolution Image Restoration Model [3] and the Context-Aware Mixed-Resolution Model [2], respectively. The programming platform comprises Python 3.7.16, PyTorch 1.3.1, CUDA 10.1.243, cuDNN 7.6.3_0, and the GPU RTX 3090, collectively enabling advanced deep learning and high-performance computations.

1) High-Resolution Image Restoration Model: The High-Resolution Image Restoration Model is designed to decrease or increase the size of images while preserving their high-resolution details. In this section, we discuss the dataset used for training and testing this model and the results obtained from testing the model on a diverse range of data.

In this model, the dataset is categorized into two types: the training dataset and the testing dataset. The training dataset includes the DIV2K dataset, which consists of 800 images showcasing various subjects, such as people, animals, buildings, flowers, landscapes, and nature. On the other hand, the testing dataset comprises three sets of images. The Set5 dataset contains five images featuring babies, butterflies, birds, human faces, and women. The Set14 dataset has 14 images showcasing subjects like people, animals, and flowers. Lastly, the B100 dataset contains 100 images with diverse topics including people, landscapes, nature, and trees.

The model was trained and tested using the DIV2K dataset. The purpose of the model was to reduce and enlarge image sizes with scaling factors of $2\times$ and $4\times$. The training dataset comprised 800 diverse images. During training, the images were divided into patches of size 192×192 for $4\times$ scaling and 96×96 for $2\times$ scaling. A batch size of 16 and a learning rate of 0.0001 were used, with the learning rate being decreased if there was no improvement in validation performance after 100 epochs. The Adam optimizer was chosen for optimization. For the loss function, the L1 norm was employed instead of the commonly used MSE (Mean Squared Error) or L2 norm. The decision to use the L1 norm was based on its ability to produce higher-quality restored images and faster training. Further details and equations can be found in references [4] and [5] for more in-depth information on the model and its training process.

In this model, tests on three different datasets mentioned above are performed. Each dataset had two variations: a $2\times$ reduction and a $4\times$ reduction in size. The evaluation metrics were divided into two parts: the effectiveness of image restoration measured by PSNR/SSIM and the file size reduction measured by the compression ratio. The results of these metrics were averaged for each dataset and are presented in Table I. As the image size decreased, the restored image's performance decreased while the compression ratio increased. This trade-off meant higher compression ratios led to a loss in image quality. Additionally, the experiments showed that the $2\times$ reduction produced better-restored images than the $4\times$ reduction, with a notable difference of 4-5 dB within the same dataset. Therefore, employing higher reduction levels in this model resulted in a more significant loss of image details.

2) The context-aware mixed-scale ensemble model: This hybrid context-aware image compression and restoration model can compress images and restore them with great detail. In this section, we describe the dataset used for training and testing the model and the results obtained from testing the model on a diverse range of data.

TABLE I
THE RESULTS OF TESTING ALL DATASETS USING THE HIGH-RESOLUTION IMAGE RESTORATION MODEL

Evaluation	Set5(×2)	Set5(×4)	Set14(×2)	Set14(×4)	B100(×2)	B100(×4)
PSNR/SSIM [dB/-]	38.96/0.9643	34.17/0.9196	35.84/0.9394	30.61/0.8427	33.87/0.922	29.48/0.809
Original filesize [Bytes]	166,444	166,444	351,702	351,702	281,300	281,300
Downscaled filesize [Bytes]	48,418	14,140	94,228	26,787	69,224	19,255
Compression ratio	3.44 (70.91%)	11.77 (91.50%)	3.73 (73.21%)	13.13 (92.38%)	4.06 (75.39%)	14.61 (93.15%)

TABLE II
TEST RESULTS OF ALL DATASETS WITH THE MIXED-CONTEXT ENCODER-DECODER MODEL

Evaluation	Kodak(9)	Kodak(8)	Kodak(7)	CLIC(9)	CLIC(8)	CLIC(7)
PSNR/SSIM [dB/-]	40.39/0.9908	38.97/0.9872	37.41/0.9817	40.67/0.9913	39.38/0.9876	38.12/0.9829
Original filesize [Bytes]	641,429	641,429	641,429	3,289,162	3,289,162	3,289,162
Downscaled filesize [Bytes]	78,280	62,479	48,875	321,321	243,191	183,384
Compression ratio	8.19(87.80%)	10.27(90.26%)	13.12(92.38%)	10.24(90.23%)	13.53(92.61%)	17.94(94.42%)

TABLE III
TEST RESULTS FOR THE KODAK DATASET USING THE COMBINED MODEL

Evaluation	Kodak (×2,9)	Kodak (×2,8)	Kodak (×2,7)	Kodak (×4,9)	Kodak (×4,8)	Kodak (×4,7)
PSNR/SSIM [dB/-]	33.10/ 0.9219	32.80/ 0.9111	32.47/ 0.8957	31.76/ 0.8444	31.58/ 0.8324	31.33/ 0.8136
Original filesize [Bytes]	641,429	641,429	641,429	641,429	641,429	641,429
Downscaled filesize [Bytes]	163,667	163,667	163,667	45,303	45,303	45,303
Compressed filesize [Bytes]	19,778	15,471	11,744	6,716	5,329	4,053
Original/Downscaled compression ratio	3.92	3.92	3.92	14.16	14.16	14.16
Downscaled/Compressed compression ratio	8.28	10.58	13.94	6.75	8.50	11.18
Original/compressed filesize compression ratio	32.43 (96.92%)	41.46 (97.59%)	54.62 (98.17%)	95.51 (98.95%)	120.37 (99.17%)	158.26 (99.37%)

This model categorizes the dataset into two types: training data and testing data. The training data consists of the YFCC100m dataset, which includes a wide range of images totaling 100 million, such as people, nature, and street scenes. The testing data comprises the Kodak dataset [6], consisting of 24 images showcasing various subjects like people, nature, buildings, and animals, as well as the CLIC professional validation dataset contains 41 images featuring multiple contents, including animals, objects, and natural scenery.

The dataset used for training the model is the YFCC100m dataset. Based on the context-aware mixed content autoencoder, the model is trained to compress images of different sizes and reconstruct high-resolution images. It is configured with nine settings, defining the quality level from 1 to 9, along with lambda values of 0.5, 0.4, 0.3, 0.2, 0.12, 0.06, 0.03, 0.017, and 0.01, respectively. Lower lambda values indicate higher quality. The experiments focus on manipulating the distortion rate by decreasing lambda values. Lower lambda values reduce compression but enhance the image reconstruction quality. The training process utilizes a subset of the YFCC100m dataset, consisting of 32,420 images. The training parameters include:

- Input image size: 256×256 patches
- Batch size: 8
- Learning rate: 0.00001
- Number of iterations: 1,000,000
- Optimizer: Adam

This model employs the KL-divergence loss function to minimize the loss value and optimize the compression model [7]. The emphasis is on the bit rate and distortion rate. In this model, each dataset was tested with quality values of 7, 8, and 9, which were identified as the optimal values for image restoration. The evaluation consisted of two aspects: the performance of image restoration measured by PSNR/SSIM, and the compression capability measured by the compression ratio. The results, shown in Table II, indicated that higher image quality led to improved image restoration performance while reducing the compression ratio. Notably, when comparing datasets in terms of image quality, the CLIC dataset exhibited significantly higher compression ratios with similar image restoration performance compared to the Kodak dataset. Furthermore, larger-sized images with repetitive or similar pixels facilitated better compression. Therefore, the proposed compression model demonstrated superior performance on datasets with larger image sizes, such as the CLIC dataset with an average size of 2000×2000

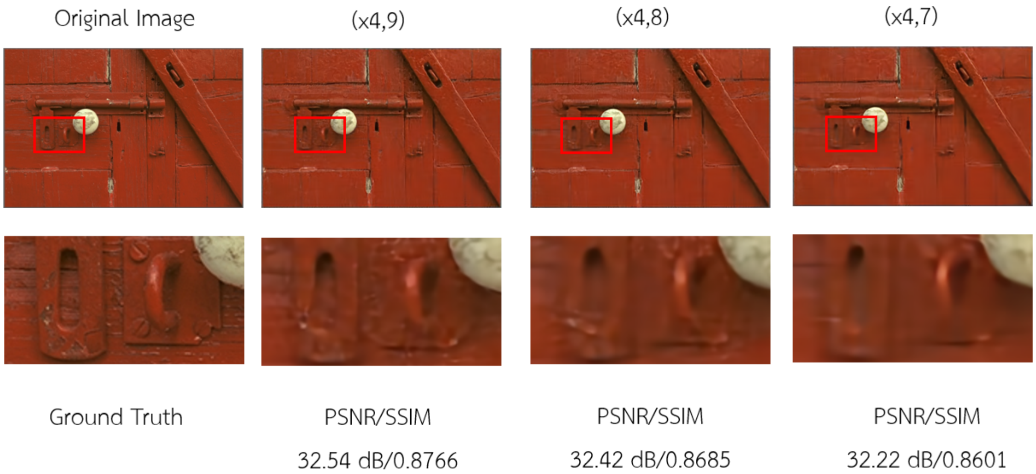

Fig. 2. Examples of image results from testing the Kodak dataset with three parameter configurations: (×4,9), (×4,8), and (×4,7).

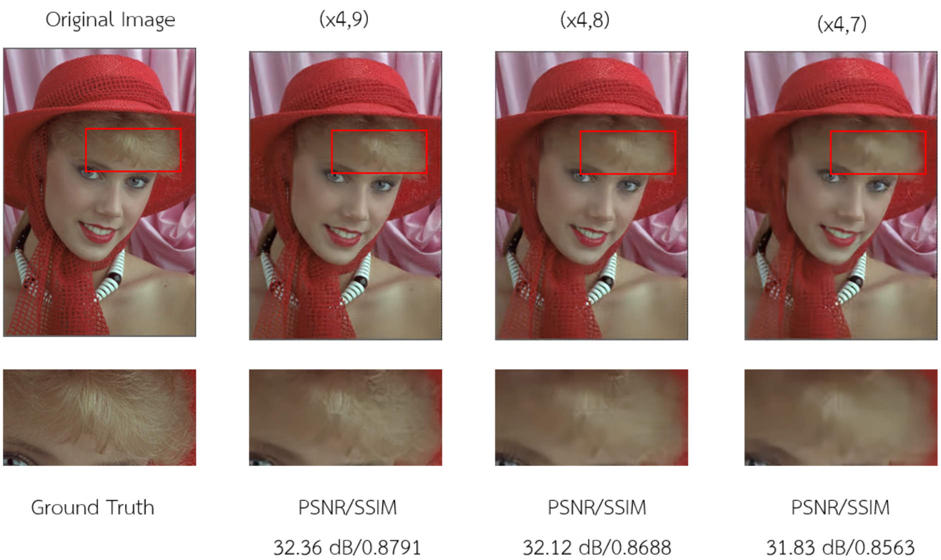

Fig. 3. Examples of image results from testing the Kodak dataset with three parameter configurations: (×4,9), (×4,8), and (×4,7).

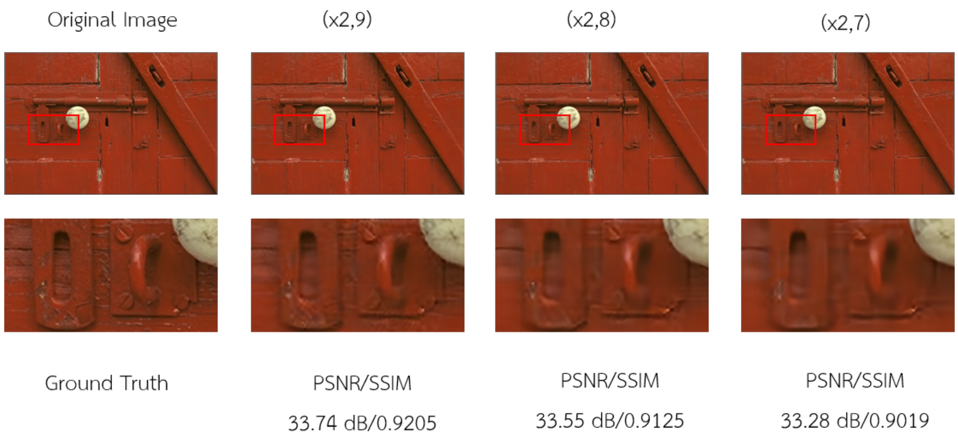

Fig. 4. Examples of image results from testing the Kodak dataset with three parameter configurations: (×2,9), (×2,8), and (×2,7).

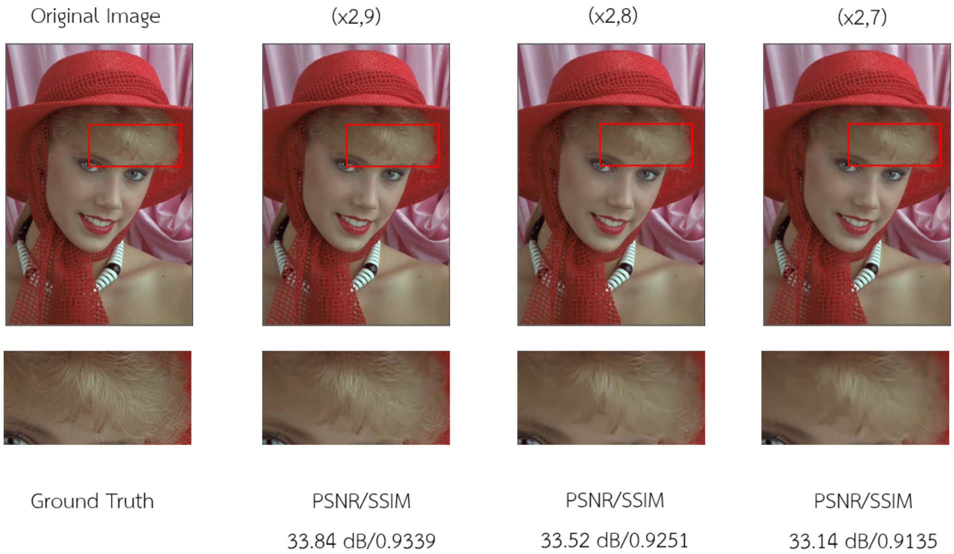

Fig. 5. Examples of image results from testing the Kodak dataset with three parameter configurations: (×2,9), (×2,8), and (×2,7).

pixels, as opposed to datasets with smaller image sizes like the Kodak dataset with a size of 768×512 pixels.

III. Experimental results of the integrated model

In this work, two models were used to achieve different goals in image processing. The first model, the mixed context entropy model [2], focused on compressing and resizing images to reduce file sizes. The second model, the high-resolution image restoration model [3], aimed to enhance image details. These two models were combined into an integrated model for the project, as shown in Figure 1.

This integrated model divides the related dataset into two categories: the training and testing datasets. The training dataset consists of two subsets: the Wave-ECC and DIV2K datasets. The Wave-ECC dataset contains 100 million images, showcasing diverse subjects such as people, nature, and street scenes. The DIV2K dataset comprises 800 images with various subjects, including people, animals, buildings, flowers, landscapes, and nature. On the other hand, the testing dataset includes the Kodak dataset, which consists of 24 diverse images featuring subjects like people, nature, buildings, and animals.

The training details of the integrated model are as follows. The training process utilizes two datasets: the Wave-ECC dataset and the DIV2K dataset. Each dataset is separately used to train the corresponding model. The first is the High-Fidelity Image Restoration model, which focuses on generating high-resolution images. The second model is the Mixed-Context Entropy-Adaptive Inpainting model. Both models are trained sequentially. The integrated model's training involves optimizing each model's loss function. These loss functions are independent of each other.

Integrated model test results were performed using the Kodak dataset. This model was tested in six different configurations and ranked in order of their ability to restore images from high to low resolution. These configurations are (x2.9), (×2.8), (×2.7), (×4.9), (×4.8), and (×4.7), where '×' represents the shrink/enlarge factor, is given a quality value. The metric is divided into two parts. Image restoration performance as measured by PSNR/SSIM, and a compression ratio of the resulting file as measured by the compression ratio. Results for these metrics were averaged for each dataset and are shown in Table III. We found that all six configurations achieved over 40% compression, and all images in the project were successfully restored. Furthermore, our experiments show that configuration (×2,9) performs best with PSNR 33.10 dB and SSIM 0.9219. Additionally, a compression ratio of up to 32.43 was achieved, resulting in a 96.92% reduction in file size.

Next, we present examples of image results obtained from testing the Kodak dataset. The top row of images represents the full-size original images and the bottom row display enlarged images so that the clarity of the restoration results obtained with different experimental parameter settings can be compared and verified.

We prepare a total of six parameter configurations(×4.9), (×4.8), (×4.7), (×2.9), (×2.8), (×2.7). These configurations correspond to two downsampling ratios, 4× and 2×, combined with three quality levels, 9, 8, and 7, respectively. By examining these different settings, we can assess the impact on the restoration outcome and visualize the trade-off between image size reduction and the level of image quality. We provide two sample images from the Kodak dataset in our illustration, each demonstrating the results achieved under a different parameter setting.

From the examples of image results obtained from testing the Kodak dataset using different compression settings, as shown in Fig.2, specifically (×4,9), (×4,8), and (×4,7), it is observed that each setting produces different outcomes. Among the three settings, (×4,9) exhibits the highest performance in image restoration. However, upon closer examination of the enlarged images, it is evident that the restored images appear significantly blurred or suffer from noticeable data loss. This observation aligns with human perception, as the generated images exhibit a PSNR value that is not significantly low, around 30 dB, which is generally acceptable. However, the SSIM value obtained is below 0.9, indicating a data loss in the restored images.

According to Fig.3, there are three settings: (×4,7), (×4,8), and (×4,9), which exhibit progressively improved performance in image restoration. However, upon observing the enlarged images, specifically the hair area of the female subject, it can be noted that the restored hairlines are hardly recognizable. When examining the setting (×4,7), it is observed that the restoration is improved, but not sufficiently compared to the (×4,9) setting. It is worth mentioning that the PSNR value of the restored images is not significantly low, exceeding 30 dB, which is generally acceptable. However, the obtained SSIM value is notably low, below 0.9, indicating a substantial data loss in the restored images.

From the examples of image results obtained from testing the Kodak dataset using different compression settings, as shown in Fig.4, specifically (×2,9), (×2,8), and (×2,7), it is evident that each setting produces different outcomes. There are three settings: (×2,7), (×2,8), and (×2,9), which exhibit progressively improved performance in image restoration. However, upon observing the enlarged images, it is clear that the best image restoration is achieved with the (×2,9) setting. This is because the details on the door handle are the most complete and precise, despite some slight blurriness. The PSNR and SSIM values for this setting are both not lower than 30 dB and 0.9, respectively. In Fig.5, it is evident that each setting produces different outcomes. There are three settings: (×2,7), (×2,8), and (×2,9), which exhibit progressively improved performance in image restoration. However, upon observing the enlarged images, particularly the hair region of the female subject, it is found that the (×2,9) setting provides the most comprehensive level of detail. The PSNR value is not significantly low, exceeding 30 dB, and the SSIM value obtained is greater than 0.9. This indicates that the restored image has relatively minimal loss compared to the original image, especially compared to the results obtained from the 4× downsampling shown in Fig.3.

IV. Discussion

After studying relevant literature, selecting and combining models to create the desired composite model has been possible. A Contextual Enhancement Generator (CEG) model and a High-Resolution Image Synthesis (HRIS) model, namely, make up the combined model. Each model has undergone experimental testing, and the findings support the given literature. A composite model from the two separate models was also tested and demonstrated using the restored photos. The research's scope reveals that the dataset used for testing, which consists of large-sized images, exceeded the capabilities of the computer's main memory, limited to 24 GB. Thus, it was not possible to conduct tests on this dataset. However, the presented composite model can be used without issues with a larger memory size.

V. Conclusion

Creating the new composite model begins with passing the high-resolution image through the HRIS model, followed by the compression model. This approach allows for high file compression rates while maintaining bitstream data using arithmetic coding. Afterward, the image is restored using the compression model and passed through the upscaling model, resulting in an image with the original desired dimensions. The tested composite model applied to the Kodak dataset shows that all six experimental settings achieve compression rates of no less than 40 percent, meeting the project's objectives and scope. The best-performing composite model is (×2,9), with an average file compression rate of 96.92%. It also successfully restores images with satisfactory performance compared to the image file size compression ratio. It achieves a PSNR value of 33.10 dB and an SSIM value of 0.9219. The generated sample image demonstrates the ability to perceive the details of the restored image with human visual perception.

References

[1] G. K. Wallace, 'The JPEG still picture compression standard'.*IEEE transactions on consumer electronics*, 38(1), xviii-xxxiv, 1992.

[2] J. Lee, S. Cho, and S.-K. Beack, 'Context-adaptive entropy model for end-to-end optimized image compression', *arXiv preprint arXiv:1809. 10452*, 2018.

[3] W. Sun and Z. Chen, 'Learned image downscaling for upscaling using content-adaptive resampler', *IEEE Transactions on Image Processing*, vol. 29, pp. 4027–4040, 2020.

[4] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, 'Loss functions for image restoration with neural networks', *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[5] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, 'Enhanced deep residual networks for single image super-resolution', in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[6] E. Kodak. Kodak lossless true color image suite (photocd pcd0992), 1993. URL http://r0k.us/graphics/kodak/.

[7] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, 'Variational image compression with a scale hyperprior', *arXiv preprint arXiv:1802. 01436*, 2018.