International Scientific Journal "Internauka" https://doi.org/10.25313/2520-2057-2025-3

Технічні науки

UDC 004.932

У Джуці

Faculty of Cyber security, Software Engineering and Computer Sciences

International Humanitarian University

JOINT LEARNED-IMAGE COMPRESSION AND SUPER RESOLUTION

Summary. *Recent advancements in learning-based image coding have demonstrated promising outcomes. These codecs utilize deep neural networks to reduce dimensionality at the stage where a linear transform would typically be applied. This signal representation, known as latent space, can be interpreted by other deep neural networks without decoding, offering benefits for various image processing tasks. In this study, we establish baselines by combining learned-image compression and Super Resolution (SR) for hybrid modeling. Specifically, we experiment with two baselines: one leveraging fixed image compression and SR models, and the other integrating fixed image compression and adaptively learned-SR models. Experimental results indicate that our second approach yields better perceptual quality than baseline 1. Moreover, our baseline 2 achieved a 23.33% improvement in BD-Rate PSNR and 0.44 dB in BD-PSNR when evaluated on the DIV2K validation set, and a 20.23% gain in BD-Rate PSNR and 0.34 dB in DB-PSNR when evaluated on the JPEGAI test set. Furthermore, we explored the impact of training data sets on the performance of image compression models to determine the optimal choice of training dataset for our hybrid modeling.*

Key words: *Learned-image compression, image coding, Super Resolution, image enhancement, image resotration.*

Introduction. Image compression is a fundamental component of signal processing and computer vision, serving as a crucial low-level image processing task. Its significance extends to various critical applications, including medical imaging, satellite imaging, multimedia services, telecommunications, the Internet of Things (IoT), and security (Viswanathan & Palanisamy, 2023). Image compression reduces the bits needed for storage and transmission. Efficient data transfer is crucial in today's Internet. Uncompressed high-resolution images consume excessive bandwidth. Therefore, compression optimizes storage and transmission, vital for applications like video conferencing and satellite imagery. Lossy methods include transform, discrete cosine, vector quantization, fractal, singular value decomposition, and wavelet coding. Lossless methods include run-length, arithmetic, Huffman, and Lempel-Ziv coding (Sandeep et al., 2023) as illustrated in Fig. 1.

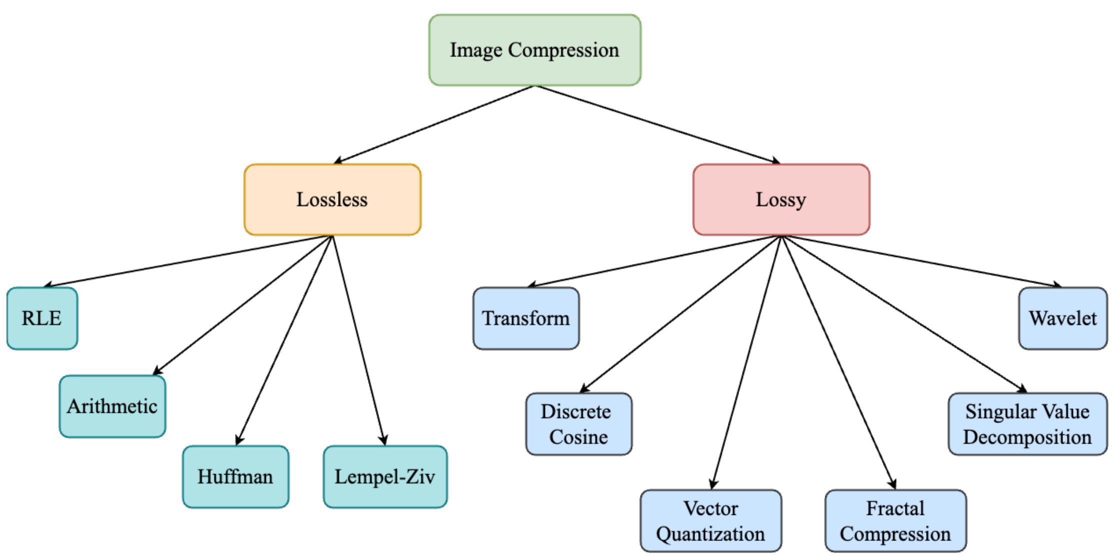


Fig. 1. Image compression techniques

Recent substantial improvements in powerful computation along with superior and wide-ranging machine learning (ML) and deep learning-based artificial neural network (ANN) methods (Fig. 2) have allowed image compression to further improve in reducing JPEG artifacts, compression perceptual quality, Peak Signal-to-Noise Ratio (PSNR), and computational complexity (IEEE, 2019). Deep learning image compression replaces linear transforms with CNNs, mapping pixels to a lower-dimensional latent space. Traditional codecs use linear transforms. (Robinson & Kecman, 2003), proposed a grayscale compression method using a joint ML-SVM and DCT, improving visual quality over JPEG (Xu, IEEE Beijing Section, & IEEE, 2018) used supervised regression and PCA for color image compression, minimizing prediction error and seed selection complexity. (Khan Gul et al., 2022) proposed RNNSC, an RNN-based stereo image compression, leveraging redundancy to reduce bit rate. Recurrent units enable adjustable compression without retraining. (Gregor et al., 2016), used a variational autoencoder for improved latent variable modeling on ImageNet and Omniglot, achieving high-quality 'conceptual compression' by retaining global information over low-level details.

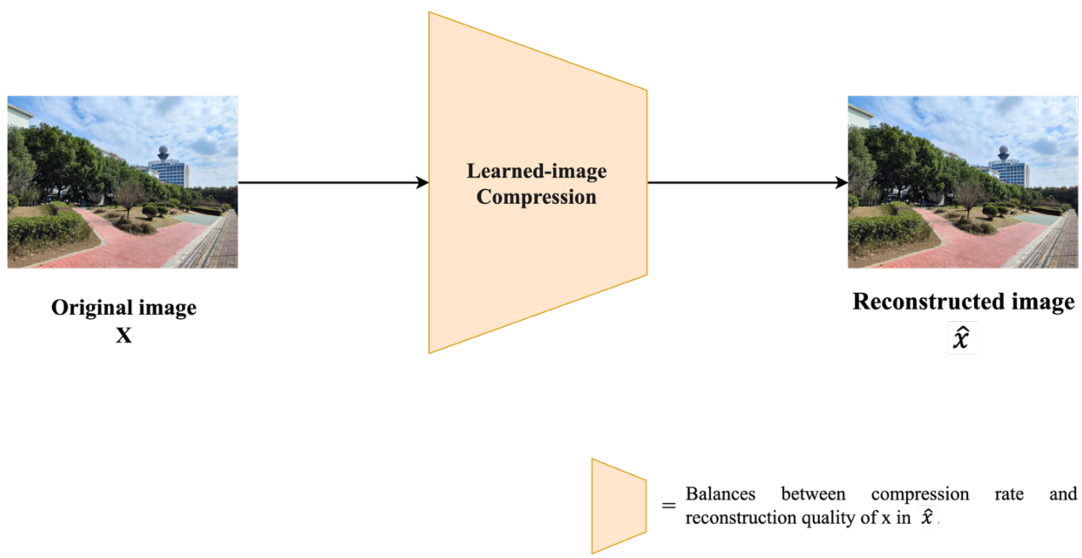


Fig. 2. Latent representation of the input image is quantized, and utilized for compression and reconstruction to obtain the final reconstructed image

(IEEE, 2017) Used a 12-layer CNN to reduce compression artifacts. "DeepN-JPEG" (Liu et al., 2018) achieved 3.5x higher compression than JPEG, maintaining quality. (IEEE, 2018)Applied a symmetric CAE with PCA for better coding efficiency. (Ballé et al., 2018) Introduced a VAE with a convoluted scale hyperprior, improving visual quality over prior methods. (Minnen, Ballé, & Toderici, 2018) extended (Ballé et al., 2018) with a hierarchical and autoregressive entropy model, outperforming BPG. SISR aims to recover HR images from LR ones. CNN approaches, starting with SRCNN (Dong et al., 2014) , have significantly improved SR, particularly PSNR (Haris, Shakhnarovich, & Ukita, 2018). However, PSNR-oriented methods often over-smooth images (Kim, Lee, & Lee, 2015). GANs (Ledig et al., 2016) address this by enhancing perceptual quality. ESRGAN (Wang et al., 2018), using RDDB blocks and relativistic GAN, won the PIRM-SR Challenge (Blau et al., 2018), emphasizing perceptual index. It employs residual scaling, smaller initialization, and enhanced perceptual loss for detailed texture recovery.

Vision Transformers (ViTs) (Chen et al., 2022) excel in Super Resolution (SR) via Multi-Head Self-Attention (MHSA), capturing long-range dependencies. However, MHSA's quadratic complexity limits inference speed. ALAN addresses this with Asymmetric Depth-Wise Convolution Attention (ADWCA), improving both SR quality and speed.

Learning-based image coding's latent space allows direct processing by other networks, benefiting SR (Chen, Qin, & Wen, 2024). This study explores SR within the compressed domain, comparing fixed compression/SR networks versus an SR network adapted for compressed images. In this investigation, we examine various approaches for applying super resolution to the output of a compression network. Specifically, we consider two approaches: (1) using a fixed compression and Super Resolution networks, and (2) an adapted version of the Super Resolution network retrained to operate on images in compressed scenarios.

Related work. In this section, we initially discuss learned-image compression, followed by a discussion on single image Super Resolution.

1. Learned-image compression

Similar to all other lossy compression techniques, machine learning methods for lossy image compression operate on a fundamental principle: an image, typically modeled as a vector of pixel intensities (x), undergoes quantization, reducing the amount of information required to store or transmit it, but also introducing error at the same time. Usually, the pixel intensities are not quantized directly. Instead, the quantization takes place in an alternative (latent) representation of the image, a vector in some other space (y), yielding a discrete-valued vector ($\hat{y}$). Therefore, it can be losslessly compressed using entropy coding methods, such as arithmetic coding.

Learning-based image methods for end-to-end coding have emerged as powerful tools in the context of image compression, and are capable in some instances of surpassing the performance of traditional approaches (Agustsson et. Al, 2017).

2. Single image super resolution

Super resolution is a category of techniques and methods to upscale raster images by a factor of two or more. Single-image super resolution focuses on a solitary image, lacking the ability to leverage correlation between subsequent frames as seen in multi-view or video super resolution. This technique represents an evolutionary step beyond traditional image re-sampling methods such as bilinear, bicubic, and Lanczos filtering, with the latter being regarded as the most effective among conventional approaches. In recent years, advancements in deep learning have enabled super resolution methods to achieve outstanding visual quality for up-scaling factors of four or higher. This subsection will examine a variety of learning-based super resolution techniques and also explain the architecture of SR network which we adopted for our work.

Photo-realistic single image super-resolution using a generative adversarial network (SRGAN) is a pioneering super-resolution model that applies GANs, incorporating deep residual networks that diverge from relying solely on Mean Square Error (MSE) as the primary optimization target (Minnen, Ballé, & Toderici, 2018). SRGAN deviates significantly from previous super-resolution methods by introducing a novel perceptual loss function based on high-level feature maps derived from the VGG network. Prior GANs, first introduced by Goodfellow in 2014, typically accept random noise as input to the generator. Conversely, SRGAN's generator accepts a lower-resolution image as input, while the discriminator operates conventionally. The primary distinction lies in the loss function, which minimizes the Euclidean distance between feature representations of reconstructed and original images obtained from the pre-trained VGG19 network. This approach yields generated images that are more faithful to a natural manifold rather than pixel-wise comparisons. Enhanced deep residual networks for single image super-resolution (EDSR) Agustsson et. Al, 2017 is a state-of-the-art super resolution residual model, securing first and second place at the NTIRE 2017 competition. It builds upon SRResNet with an improved architecture designed for faster computation and superior performance outcomes.

Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) (Chen, Qin, & Wen, 2024), illustrated in Fig. 3(a), acknowledged a limitation in the SRGAN architecture, namely the propensity to produce unrealistic visual artifacts, often referred to as hallucinations. To elevate visual quality, the authors refined the network architecture, adversarial loss, and perceptual loss. A novel

Residual-in-Residual Dense Block which is composed of three residual dense blocks with residual scaling parameter (β), envisioned in Fig. 3(b) and 3(c), without batch normalization serves as the fundamental building block for the network. Therefore, in our joint compression and super-resolution framework, we utilize the ESRGAN as the decoder component.

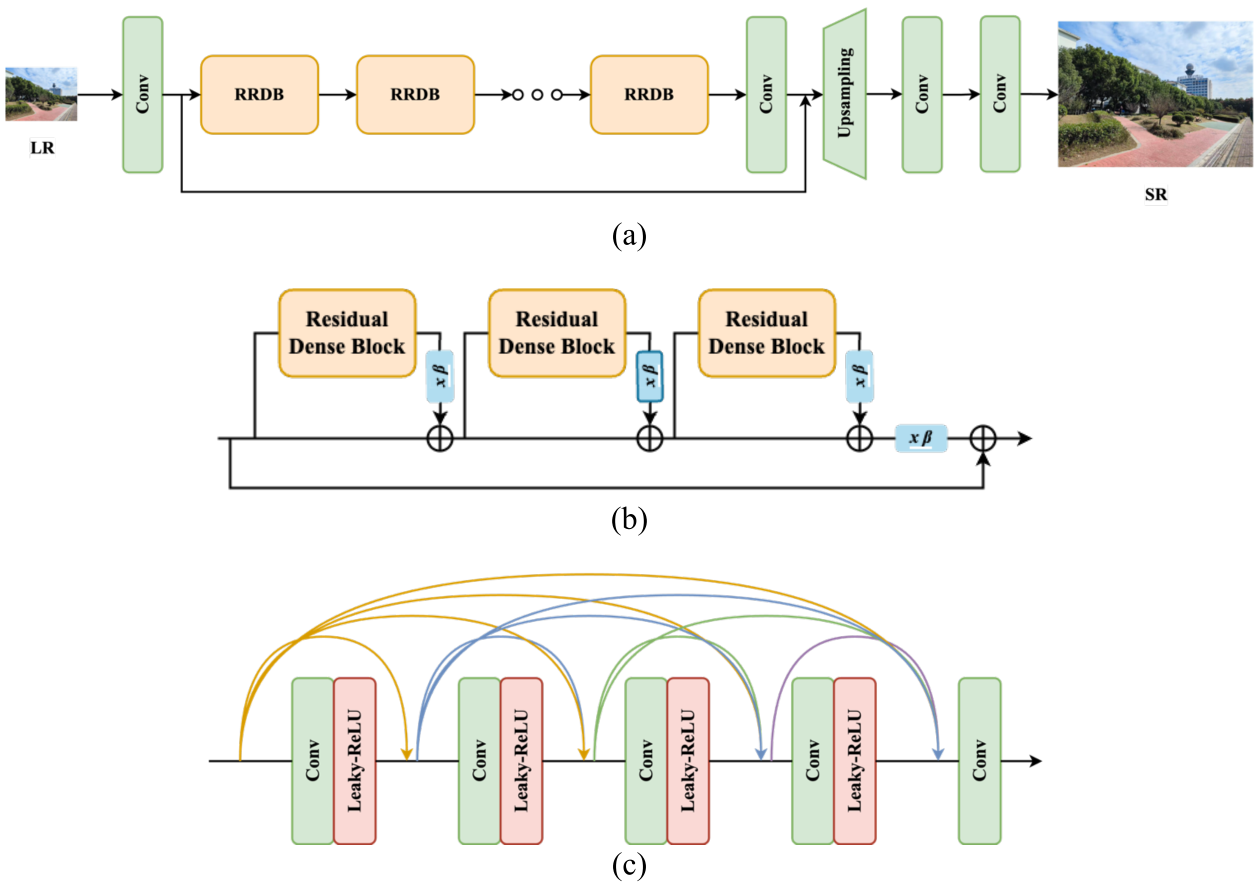


Fig. 3. Architecture of (a) enhanced super-resolution generative adversarial network (ESRGAN), (b) residual-in-residual dense block, and (c) residual dense block.

Methods

1. Utilizing fixed encoder and decoder

Lossy compression, while efficient, degrades image quality, impacting tasks like super-resolution. This study uses a fixed encoder-decoder for two-stage processing: lossy compression of the LR image, followed by 4x super-resolution upscaling for reconstruction. Fig. 4 illustrates this approach.

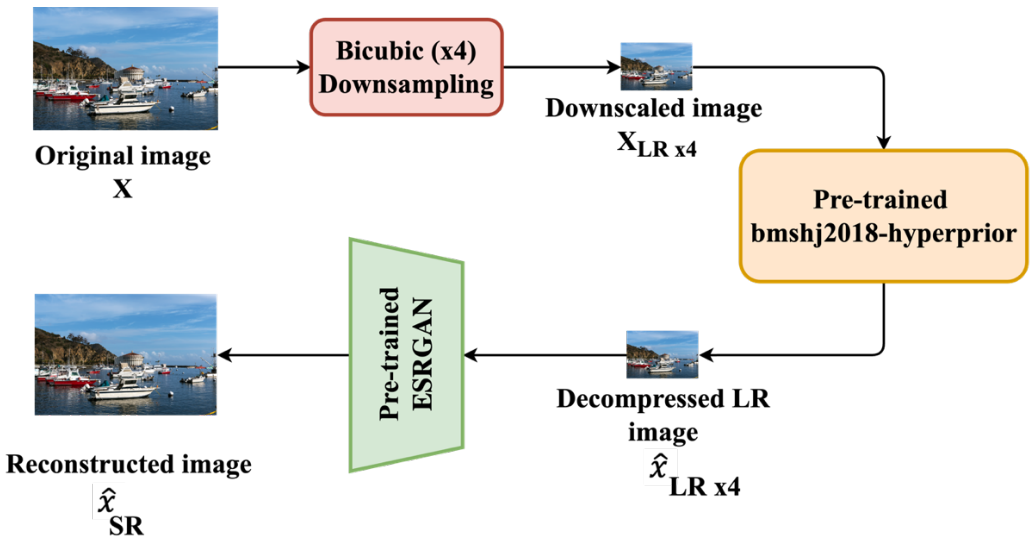


Fig. 4. Framework of baseline 1 utilizing with deep learning-based pretrained learned-image compression and super-resolution models

We utilize a priorly trained bmshj2018-hyperprior model to compress the downsampled LR image and then upscale and enhance it using a pre-trained ESRGAN. To evaluate our model's performance, we compare its output with the corresponding original image using a range of objective metrics, including Peak Signal-to-Noise Ratio (PSNR) and Multi-Scale Structural Similarity Index Measure (MS-SSIM). This fixed encoder-decoder architecture serves as our baseline 1 for joint compression and super-resolution task.

2. Utilizing fixed encoder and learned decoder

This workflow combines a pre-trained bmshj2018-hyperprior compression model with an adaptively learned SR model. The downsampled image is lossy compressed. The compression network's output (excluding entropy coding) is input to the SR network. Training follows ESRGAN, but the LR image is passed through the compression network (no entropy coding) for each compression quality. Initial compression step is the same as in our baseline 1, but the only difference is that $\hat{x}$ is then fed into the ESRGAN as input data together with its original high-resolution counterpart to train the underlying SR model to produce the final reconstructed image ($\hat{x}_{SR}$). We refer to this combination of fixed compression and learned-SR approach as our baseline 2.

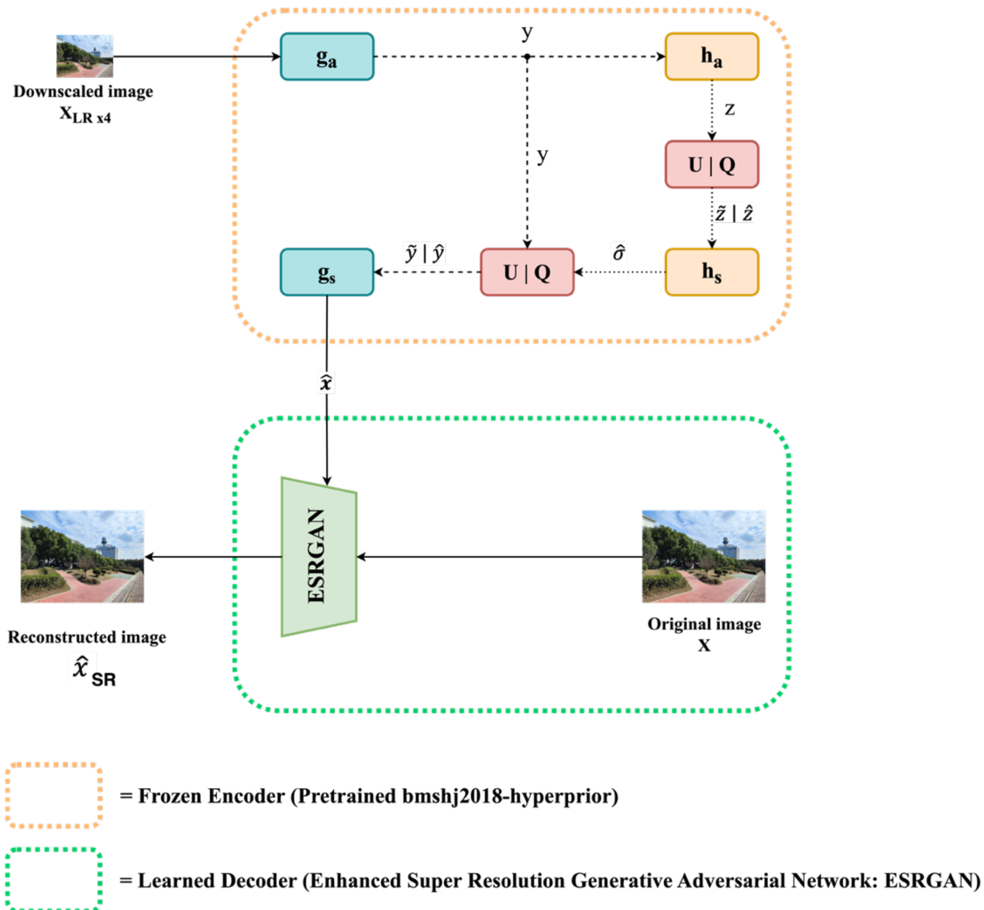


Fig. 5. Framework of baseline 2 utilizing pretrained learned-image compression and learned super-resolution models

3. Loss function

The discriminator in ESRGAN is the relativistic discriminator denoted as (D_{Ra}), which estimates the likelihood that a real image x_r appears significantly more natural compared to a fake image x_f as mentioned in equation 2 and 3. D_{Ra} is formulated as in equation (1), where $E_{x_f}[\cdot]$ is the operation of aggregating average values from all fake data within the mini-batch. σ is the sigmoid function and $C(x)$ is the non-transformed discriminator output.

$$D_{Ra}(x_r, x_f) = \sigma\left(C(x_r) - E_{x_f}[C(x_f)]\right), \quad (1)$$

$$D_{Ra}(x_r, x_f) = \sigma\left(C(x_r) - E[C(x_f)]\right) \rightarrow 1, \quad (2)$$

$$D_{Ra}(x_f, x_r) = \sigma\left(C(x_f) - E[C(x_r)]\right) \rightarrow 0, \quad (3)$$

Equation 4 and 5 define the discriminator loss and generator loss, respectively, where $x_f = G(x_i)$ and x_i refers to the input LR image and adversarial loss for generator contains both x_r and x_f .

$$L_D^{Ra} = -E_{x_r}\left[\log\left(D_{Ra}(x_r, x_f)\right)\right] - E_{x_f}\left[\log\left(1 - D_{Ra}(x_f, x_r)\right)\right], \quad (4)$$

$$L_G^{Ra} = -E_{x_r}\left[\log\left(D_{Ra}(x_r, x_f)\right)\right] - E_{x_f}\left[\log\left(D_{Ra}(x_r, x_f)\right)\right], \quad (5)$$

A new perceptual loss function, L_{percep} is introduced to the generator loss, that constrains features before activation rather than after. This approach addresses two limitations of the original design: sparse activation and inconsistent reconstructed brightness. The total loss for the generator is the sum of L_{percep}, content loss (L_1), and regularization term, equation (6).

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1, \quad (6)$$

Where $L_1 = E_{x_i}\|G(x_i) - y\|_1$, is the content loss that measures the 1-norm distance between the recovered image $G(x_i)$ and the ground-truth y, while λ and η serve as coefficients to regulate the relative importance of distinct loss components.

Experiments. To compare the compression and image enhancement performance of our proposed methods, we conducted a number of experiments using PyTorch framework.

1. Experimental tools

Anaconda Navigator simplifies experimental setups, especially for Python 3.8 and package management. It provides a user-friendly interface for consistent environment creation, package installation, and dependency management using conda. Environments isolate projects, crucial for multi-faceted research. Anaconda Navigator also facilitates access to libraries like TensorFlow, PyTorch, Pandas, and NumPy. Python's versatility, libraries, and community make it essential for deep learning (DL) and computer vision (CV). Its simplicity enables rapid prototyping, and libraries support complex tasks like image recognition and NLP. Python is pivotal for AI advancements. PyTorch was crucial for our low-level CV research. Its dynamic neural network library offered flexibility and efficiency. The torch.autograd module automated gradient computation, streamlining training. Network building blocks (nn.Conv2d, nn.ReLU, nn.MaxPool2d) enabled modular model construction. The dynamic computation graph allowed agile model design, and GPU acceleration sped up computations. PyTorch empowered us to explore innovative methodologies in computer vision.

2. Training Details

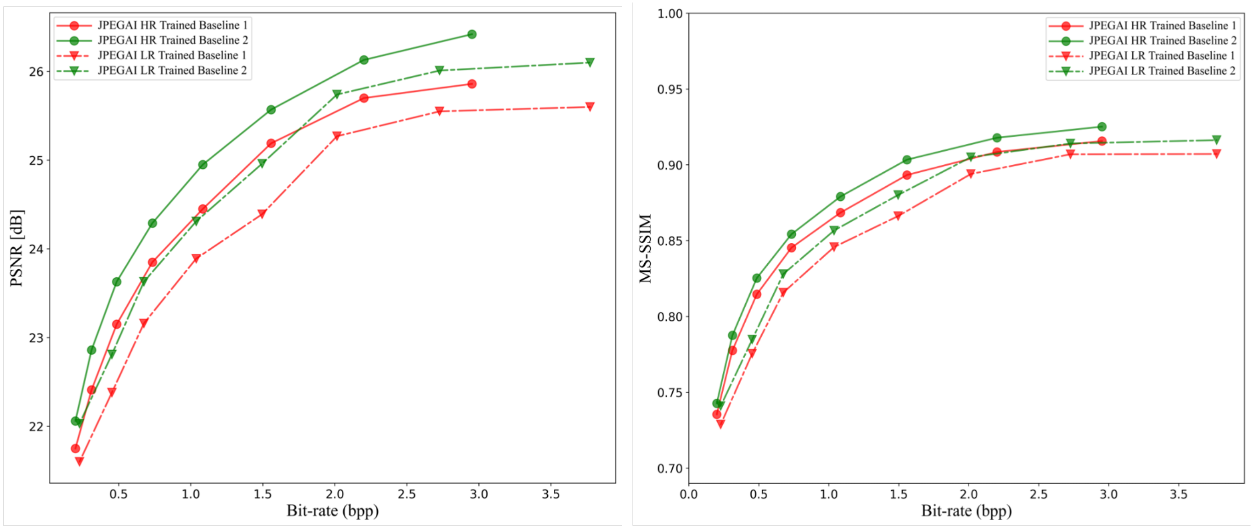
For this experiment, we employ two types of datasets: JPEGAI (https://jpegai.github.io/3-datasets/) and DF2K, which we use to train the decoder part in our baseline 2. IJPEGAI is used to train the encoder component of our baselines. bmshj2018-hyperprior models are trained for compression qualities 1-8 using varying λ. 16 encoders are trained with MSE distortion, 32 minibatch size, Adam optimizer, and 1e-4 learning rate. High-resolution training uses 256x256 patches, low-resolution uses 96x96. Trained encoders are evaluated on DIV2K, JPEGAI, and Kodak. Baseline 1 uses a pre-trained ESRGAN decoder. Baseline 2 trains ESRGAN with DF2K using pre-trained encoders. Baselines are evaluated on DIV2K and JPEGAI.

3. Experimental results

This section analyzes our joint compression and super-resolution results. Baseline 2 (pretrained compression encoder, ESRGAN-trained decoder) outperformed baseline 1 (pretrained compression/SR models) in bpp, PSNR, and MS-SSIM. We explore optimal trade-offs and discuss results, including benefits, limitations, objective/subjective evaluations, and applications. This analysis details the benchmarking results of the baseline models through objective and subjective comparisons of joint compression and super-resolution.

3.1 Objective evaluation of baselines

Fig.s 6 (a) and (b) display RD curves representing the average PSNR and MS-SSIM metrics for 100 images from the DIV2K validation dataset compressed at various bitrates. The solid lines represent baselines trained on high-resolution image datasets, specifically JPEGAI, while dashed lines indicate those utilizing an encoder trained on low-resolution images. Results on DIV2K show adaptive learning decoder (joint compression/SR) outperforms pre-trained models. High-resolution JPEGAI encoder yields best performance; low-resolution encoder (baseline 1) performs worst. Low-resolution trained models require higher bitrates for similar performance.



(a) (b)

Fig. 6. Evaluation of baseline 1 and 2 on DIV2K validation set

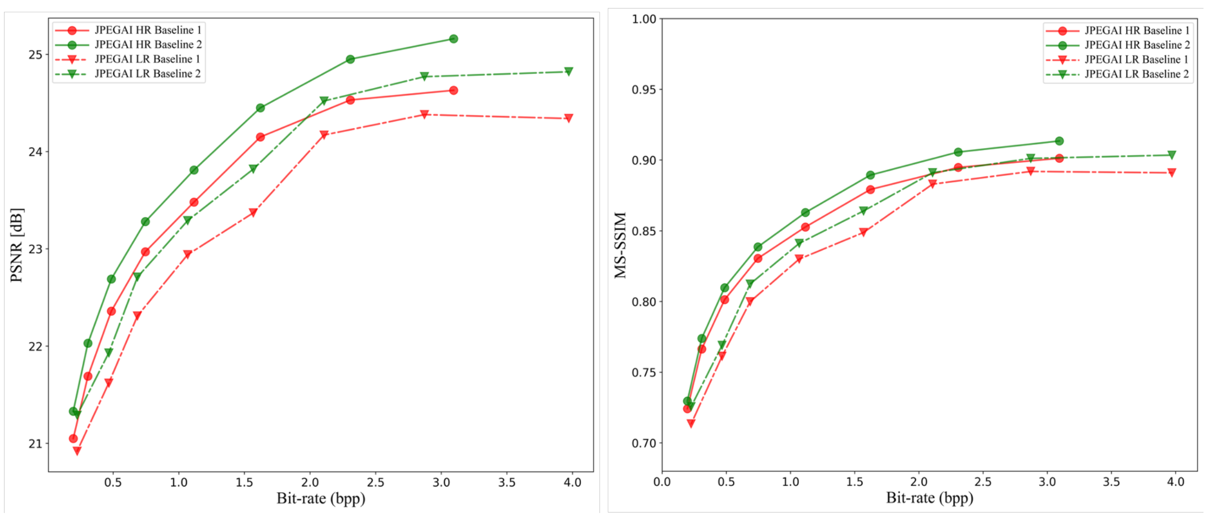
Table 1

Average difference in bitrate, PSNR, and MS-SSIM between RD curves of baselines evaluated on the DIV2K validation set. The first and second row measures the BD-Rate PSNR (%), BD-PSNR [dB], BD-Rate MS-SSIM (%), and BD-MS-SSIM between the two baselines whose encoders are trained with JPEGAI HR and LR training set, respectively.

Encoder Training Set	BD-Rate PSNR (%)	BD-PSNR [dB]	BD-Rate MS-SSIM (%)	BD-MS-SSIM
JPEGAI HR	-23.33%	0.442434177	-12.98%	-0.12975399
JPEGAI LR	-23.48%	0.461000192	-14.04%	-0.140433824

Table 1 compares bitrates, PSNR, and MS-SSIM for two baselines on DIV2K, using high- and low-resolution trained encoders. Baseline 2, with a high-resolution JPEGAI-trained encoder, significantly outperformed baseline 1. It showed a 23.33% BD-PSNR improvement (0.44 dB PSNR increase) and a 12.98% BD-Rate MS-SSIM improvement (0.129 MS-SSIM increase). With a low-resolution encoder, baseline 2 also excelled, achieving a 23.48% BD-PSNR improvement (0.46 dB PSNR increase) and a 14.04% BD-Rate MS-SSIM improvement (0.14 MS-SSIM increase).

On the JPEGAI test set, RD curves (Fig. 7 a, b) show similar performance trends. The dynamic learning decoder approach (baseline 2) outperformed pre-trained models (baseline 1). Baseline 2, with a high-definition JPEGAI-trained encoder, performed best. Baseline 1, with a low-resolution JPEGAI-trained encoder, performed worst.



(a) (b)

Fig. 7. Evaluation of baseline 1 and 2 on JPEGAI test set

Table 2

Average difference in bitrate, PSNR, and MS-SSIM between RD curves of baselines evaluated on the JPEGAI test set. The first and second row measures the BD-Rate PSNR (%), BD-PSNR [dB], BD-Rate MS-SSIM (%), and BD-MS-SSIM between the two baselines whose encoders are trained with JPEGAI HR and LR training sets, respectively

Encoder Training Set	BD-Rate PSNR (%)	BD-PSNR [dB]	BD-Rate MSSSIM (%)	BD-MSSSIM
JPEGAI HR	-20.23%	0.34144768	-11.24%	-0.112440499
JPEGAI LR	-20.89%	0.369938883	-13.47%	-0.134702901

Table 2 compares bitrates, PSNR, and MS-SSIM for two baselines on the JPEGAI test set, using high- and low-resolution trained encoders. Baseline 2, with a high-resolution JPEGAI-trained encoder, outperformed baseline 1. It showed a 20.33% BD-PSNR improvement (0.34 dB PSNR increase) and an 11.24% BD-Rate MS-SSIM improvement (0.112 MS-SSIM increase). With a low-resolution encoder, baseline 2 also excelled, achieving a 20.89% BD-PSNR improvement (0.37 dB PSNR increase) and a 13.47% BD-Rate MS-SSIM improvement (0.13 MS-SSIM increase).

3.2 Subjective evaluation of baselines

This section presents visual comparisons between two baselines that jointly compress and super-resolve images from both DIV2K and JPEGAI datasets across various bitrates. The visualizations begin with the original ground truth image, followed by a cropped portion of this image in the first column. Subsequent columns display the jointly compressed and super-resolved results at different compression qualities: 1, 4, and 8.

In the following Fig.s, we present visualizations of the performance of baselines 1 (B1) and 2 (B2), each equipped with either HR JPEGAI-trained encoders (denoted as HR B1 and HR B2) or low-resolution counterparts (denoted as LR B1 and LR B2). The labels "x4" indicate a 4-fold upsampling using a selected SR model, while "Q" represents the compression quality. For example, HR B1 x4 @ Q1 refers to the reconstructed image of baseline 1 whose encoder is trained with JPEGAI HR images, compressed at a compression quality of 1, and then upscaled by a factor of 4. The Fig.s illustrate the performance of baselines 1 and 2 by showcasing the best and second-best PSNR and MS-SSIM values for each reconstructed image at the same compression quality in red and blue colors, respectively, to facilitate visual comparison between the two baselines.

Fig. 8 presents a qualitative comparison of baselines equipped with encoders trained on high-resolution JPEGAI images, applied to the "0801.png" image from the DIV2K dataset. The results reveal that baseline 2 consistently outperformed baseline 1 in terms of PSNR across all compression quality levels (Q1, Q4, and Q8). Meanwhile, baseline 1 achieved higher MS-SSIM values than baseline 2 at compression qualities 4 and 8. This pattern is similarly observed for LR B1 and B2 with the difference that LR B1 outperforms LR B2 in both PSNR and MS-SSIM at Q-value 8, Fig. 8. In Fig. 9 and 10, we can see HR and LR B2 gave higher PSNR and MS-SSIM, except for LR B2 at quality 8. In Fig. 16 to 19, it is observed the significant performance of B2 over B1 when evaluated on JPEGAI test set.

We get clearer reconstructed images as the compression quality (Q) gets larger as we can see in Fig. 8 - 10. However, we can observe visually unpleasant artifacts in the output images when using the joint compression and super-resolution method with fixed encoder and decoder. On the other hand, when employing the method which combines frozen encoder and learned-decoder returns visually more pleasant reconstruction with less noises. The results indicates that when we evaluate the baselines on DIV2K and JPEGAI images, the baseline which has learned-SR model is able to generate more perceptually pleasing images with fewer artifacts.

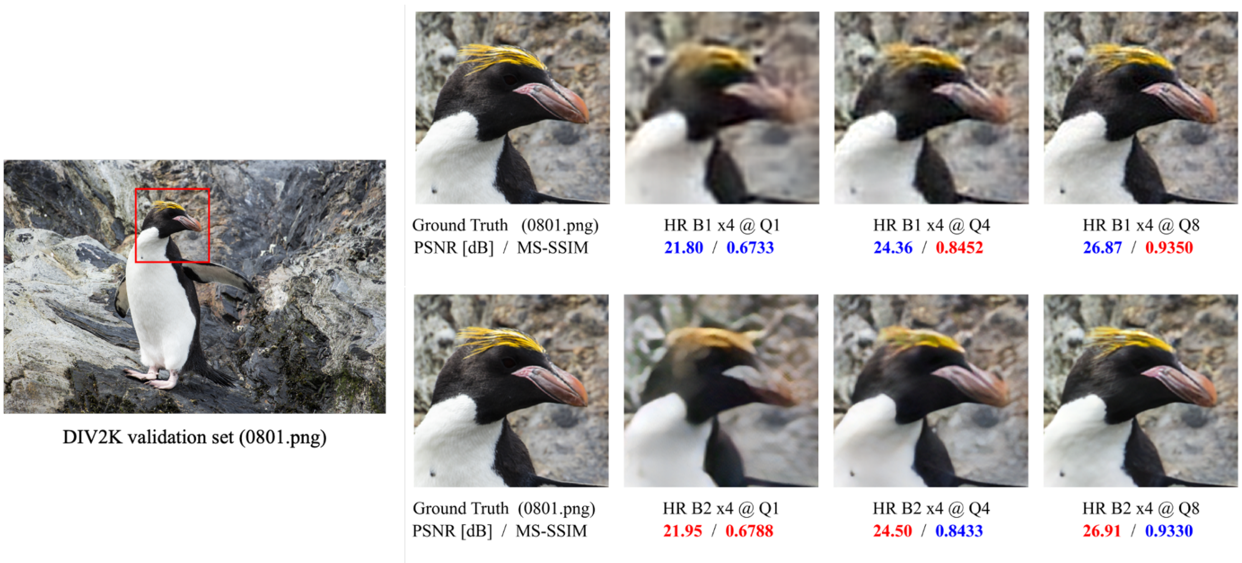


Fig. 8. Perceptual quality comparison (best and second best) between baselines utilizing the encoder trained with JPEGAI high-resolution images (DIV2K validation set image 0801.png)

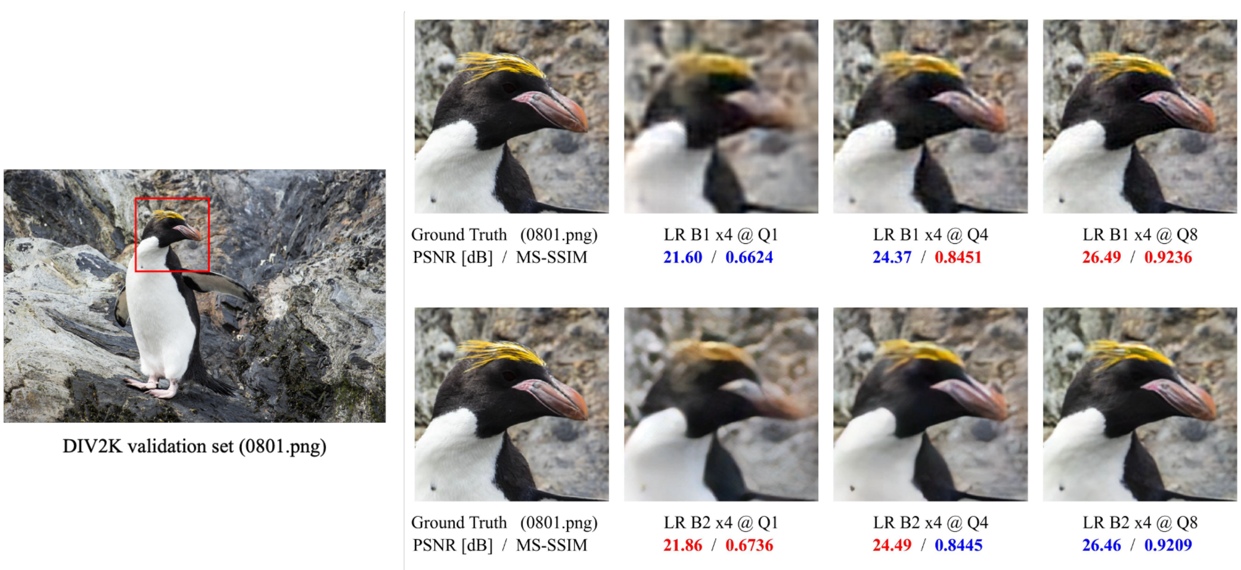


Fig. 9. Perceptual quality comparison (best and second best) between baselines utilizing the encoder trained with JPEGAI low-resolution images (DIV2K validation set image 0801.png)

Fig. 10. Perceptual quality comparison (best and second best) between baselines utilizing the encoder trained with JPEGAI high-resolution images (DIV2K Validation Set Image 0881.png)

Discussion. Previous results showed our baselines performed SR in the compression domain, but improvements are needed. Baseline 1 used pre-trained encoders/decoders; baseline 2 used a pre-trained encoder and learned decoder. Future work involves developing an end-to-end trainable network, requiring optimal dataset selection and a method to directly map latent representations for SR upsampling. This section examines DF2K dataset training impact on encoder performance via RD curves on DIV2K, JPEGAI, and Kodak.

Conclusion. This study explored two hybrid models for deep learning image compression and super-resolution: pre-trained models and a fixed encoder with learned SR. Evaluations on DIV2K and JPEGAI showed the learned SR approach outperformed the pre-trained model, both subjectively and objectively. Pre-trained model reconstructions exhibited compression artifacts. Training dataset impact on compression was also investigated. Future work will focus on end-to-end trainable joint compression and super-resolution methods.

References

1. P. Viswanathan and K. Palanisamy, "Predictive Codec of Medical Image Compression Using Subb and Thresholding," in Proceedings of the 2023 International Conference on Intelligent Systems for Communication, IoT and Security, ICISCoIS 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 574–579. doi: 10.1109/ICISCoIS56541.2023.10100453.

2. P. Sandeep, K. N. Reddy, N. R. Teja, G. K. Reddy, and S. Kavitha, "Advancements in Image Compression Techniques: A Comprehensive Review," in Proceedings of the 2nd International Conference on Edge Computing and Applications, ICECAA 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 821–826. doi: 10.1109/ICECAA58104.2023.10212295.

3. 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE.

4. J. Robinson and V. Kecman, "Combining support vector machine learning with the discrete cosine transform in image compression," IEEE Trans Neural Netw, vol. 14, no. 4, pp. 950–958, Jul. 2003, doi: 10.1109/TNN.2003.813842.

5. B. Xu, Institute of Electrical and Electronics Engineers. Beijing Section, and Institute of Electrical and Electronics Engineers, Proceedings of 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC 2018) : May 25-27, 2018, Xi'an China.

6. M. S. Khan Gul, H. Suleman, M. Batz, and J. Keinert, "RNNSC: Recurrent Neural Network-Based Stereo Compression Using Image and State Warping," Institute of Electrical and Electronics Engineers (IEEE), Jul. 2022, pp. 455–455. doi: 10.1109/dcc52660.2022.00066.

7. K. Gregor, F. Besse, D. J. Rezende, I. Danihelka, and D. Wierstra, "Towards Conceptual Compression," Apr. 2016, [Online]. Available: http://arxiv.org/abs/1604.08772 (date of access: 15.03.2025).

8. IEEE Computational Intelligence Society, International Neural Network Society, and Institute of Electrical and Electronics Engineers, IJCNN 2017 : the International Joint Conference on Neural Networks.

9. Z. Liu et al., "DeepN-JPEG: A deep neural network favorable JPEG-based image compression framework," in Proceedings - Design Automation Conference, Institute of Electrical and Electronics Engineers Inc., Jun. 2018. doi: 10.1145/3195970.3196022.

10. Institute of Electrical and Electronics Engineers, PCS 2018 : 2018 Picture Coding Symposium (PCS) : proceedings : 24-27 June 2018, San Francisco, California, USA.

11. J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, "Variational image compression with a scale hyperprior," Jan. 2018, [Online]. Available: http://arxiv.org/abs/1802.01436 (date of access: 15.03.2025).

12. D. Minnen, J. Ballé, and G. Toderici, "Joint Autoregressive and Hierarchical Priors for Learned Image Compression," Sep. 2018, [Online]. Available: http://arxiv.org/abs/1809.02736 (date of access: 15.03.2025).

13. C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1501.00092 (date of access: 15.03.2025).

14. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep Back-Projection Networks For Super-Resolution," Mar. 2018, [Online]. Available: http://arxiv.org/abs/1803.02735 (date of access: 15.03.2025).

15. C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.04802 (date of access: 15.03.2025).

16.J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," Nov. 2015, [Online]. Available: http://arxiv.org/abs/1511.04587 (date of access: 15.03.2025).

17.X. Wang et al., "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," Sep. 2018, [Online]. Available: http://arxiv.org/abs/1809.00219 (date of access: 15.03.2025).

18.K. Karwowska and D. Wierzbicki, "MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images," IEEE J Sel Top Appl Earth Obs Remote Sens, vol. 16, pp. 9886–9906, 2023, doi: 10.1109/JSTARS.2023.3322642.

19.Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM Challenge on Perceptual Image Super-resolution," Sep. 2018, [Online]. Available: http://arxiv.org/abs/1809.07517 (date of access: 15.03.2025).

20.X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating More Pixels in Image Super-Resolution Transformer," May 2022, [Online]. Available: http://arxiv.org/abs/2205.04437 (date of access: 15.03.2025).

21.Y. Ji, P. Jiang, J. Shi, Y. Guo, R. Zhang, and F. Wang, "INFORMATION-GROWTH SWIN TRANSFORMER NETWORK FOR IMAGE SUPER-RESOLUTION," in Proceedings - International Conference on Image Processing, ICIP, IEEE Computer Society, 2022, pp. 3993–3997. doi: 10.1109/ICIP46576.2022.9897359.

22.Q. Chen, J. Qin, and W. Wen, "ALAN: Self-Attention Is Not All You Need for Image Super-Resolution," IEEE Signal Process Lett, vol. 31, pp. 11–15, 2024, doi: 10.1109/LSP.2023.3337726.

23.G. Toderici et al., "Variable Rate Image Compression with Recurrent Neural Networks," Nov. 2015, [Online]. Available: http://arxiv.org/abs/1511.06085

24. G. Toderici et al., "Full Resolution Image Compression with Recurrent Neural Networks," Aug. 2016, [Online]. Available: http://arxiv.org/abs/1608.05148 (date of access: 15.03.2025).

25. R. Timofte et al., "NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE Computer Society, Aug. 2017, pp. 1110–1121. doi: 10.1109/CVPRW.2017.149.

26. E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE Computer Society, Aug. 2017, pp. 1122–1131. doi: 10.1109/CVPRW.2017.150.