A Novel Deep Learning-Based Compressed Image Enhancement Method for Machine Consumption

Zhenchao Ma, Forrest Paul, Sara Zhang, Hongjing Chen, Hamid Reza Tohidypour, and Panos Nasiopoulos
Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver, Canada
zhenchaoma@ece.ubc.ca, {linus744, xiyuzhan}@student.ubc.ca, {hjchen21, htohidyp, panosn}@ece.ubc.ca

***Abstract*—Image compression reduces storage, and transmission demands but often degrades image quality, introducing artifacts such as blurring and blocking. While deep learning-based methods have shown remarkable progress in the enhancement of compressed images, most of these approaches are designed with human perception in mind, focusing on improving subjective visual quality. As the field of artificial intelligence continues to evolve, the consumption of images by machines, rather than humans, has become increasingly relevant. Compressed images, when fed into machine learning models, can cause significant performance degradation due to distortions introduced during compression. To address this gap, we propose a joint restoration-classification network specifically designed to enhance compressed images for machine consumption. Our approach combines an image enhancement network with an image classification network, using a linear combination of Charbonnier and cross-entropy loss terms to optimize classification accuracy while balancing restoration metrics such as PSNR and SSIM. Our experiments show that our approach increases top-1 classification accuracy by 6.2% for JPEG compressed images at quality level 40 and by 12.2% for images at quality level 10, compared to the baseline performance on the same compressed images without enhancement.**

***Keywords*—Image enhancement, Image compression, Machine Consumption, Swin Transformer**

I. Introduction

Image compression is vital for reducing storage and transmission needs in applications like cloud storage services, remote sensing, and augmented reality. However, it often involves a trade-off between file size and image quality, leading to distortions such as blurring, blocking artifacts, and loss of fine details. These issues are particularly problematic in areas requiring high-quality visuals, such as medical imaging, surveillance, and professional photography. To address these issues, deep learning-based approaches have emerged as a powerful solution for image quality enhancement [1], [2], [3]. These methods have shown remarkable progress in the enhancement of compressed images, leveraging neural networks to restore fine details and reduce artifacts. However, most of these methods are designed with human perception in mind, aiming to improve the subjective visual quality of images.

As the field of artificial intelligence continues to evolve, the use of images by machines has increased. Many modern applications such as object detection [4], image classification [5], and autonomous driving systems [6] rely on visual data to perform critical tasks. The performance of machine learning models in these visual tasks is significantly influenced by the quality of the input images. Compressed images, when fed into these models, can result in degraded performance due to the inherent distortions introduced by compression. Research has shown that models like DenseNet [7] and ResNet [8], which excel on high-quality datasets, perform poorly on low-quality images [9]. Other studies [10][11] have demonstrated that image distortions, such as noise, blur, and compression, significantly reduce the accuracy of deep learning models, with some models experiencing over 80% accuracy loss due to minor image modifications. Despite the rapid growth of machine consumption in visual tasks, current image enhancement methods are not optimized for machine-based analysis. Instead, they focus solely on improving images for human viewing, potentially overlooking the specific quality factors that impact machine learning model performance. This misalignment between enhancement methods and machine vision needs motivates the development of new strategies aimed at improving compressed images for machine consumption. Most existing methods focus on mitigating the impact of image degradation after it has occurred, rather than enhancing compressed images for machine learning tasks directly. For instance, Liu et al. [12] explored pre-training techniques for handling adverse environmental conditions, but their work focuses on pre-existing models rather than developing dedicated enhancement techniques for compressed images.

To address these gaps, we propose a novel joint restoration-classification network specifically designed to enhance compressed images for machine-based tasks. Our approach integrates an image enhancement network with an image classification network, leveraging a linear combination of Charbonnier and cross-entropy loss terms to optimize both image restoration quality and classification accuracy. By restoring compressed images through SwinIR [13] while simultaneously classifying them with EfficientNet [14], our network improves the visual quality of compressed images and enhances their suitability for deep learning models. Our experiments on the ImageNet-1k dataset reveal significant improvements in classification accuracy. Specifically, we achieved a 6.2% increase in top-1 accuracy for JPEG compressed images at QP = 40 and a 12.2% increase for images at QP = 10, compared to the baseline performance on the same compressed images without enhancement. These results demonstrate that our joint network effectively mitigates the adverse effects of compression, allowing deep learning models to perform better on compressed images without compromising restoration quality.

979-8-3315-2096-0/25/$31.00 ©2025 IEEE

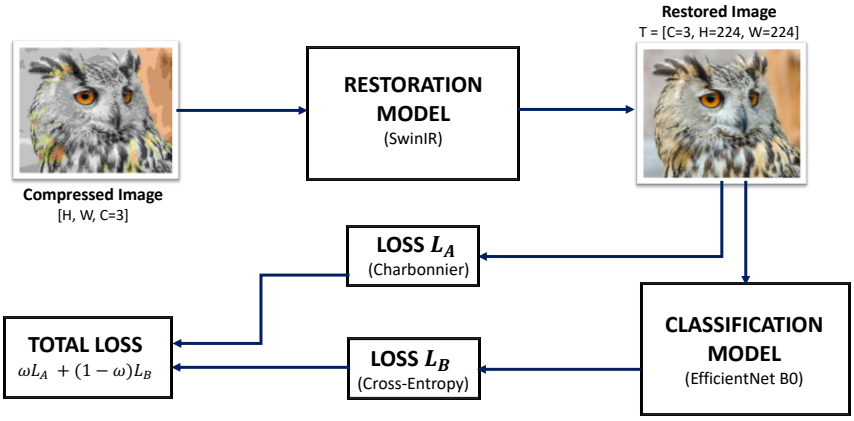

Fig. 1. The architecture of our joint restoration-classification network for machine comsumption.

The rest of this paper is organized as follows. Section II describes the overall architecture of our solution, providing an in-depth explanation of the model's structure and workflow. Section III presents the evaluation results, while Section IV concludes the paper.

II. Proposed Methods

We evaluated two state-of-the-art image enhancement techniques, SwinIR and SFNet [15], to determine the best networks for our application. Preliminary evaluations showed that SwinIR outperformed SFNet, leading us to select SwinIR for our approach. To improve the classification accuracy of compressed images, we designed and implemented a joint restoration-classification network using a SwinIR model for image restoration and a frozen EfficientNet model for image classification. This network was trained via loss feedback from both restoration and classification models. In this manner, the joint network favors the production of restored images maximally suited for classification tasks. The creation and assessment of the joint restoration-classification network can be broken into 2 core components: data preparation and network details, which will be explored in turn.

A. Data Preparation

The joint network was trained using 10 classes from the ImageNet-1k dataset, which includes a total of 12,890 training images, 500 validation images, and 100 testing images. Due to the varying resolution and aspect ratios of these images, EfficientNet and other classification networks require the images to be resized and cropped to 224x224 pixels color patches. This preprocessing step is essential before the images can be input into the network.

The data expectations for both SwinIR and EfficientNet networks differ, necessitating a series of image transformations applied between network segments. The full set of image transforms is as follows: first, the image is resized to 256 pixels on the shortest side. Next, the image is center-cropped to 224x224 pixels. For training, the image is flipped horizontally with a 50% probability, while no transformations are applied during validation or testing. The image is then converted to a tensor, with image channels swapped from [H, W, C] to [C, H, W], and pixel values normalized between [0, 1]. After image restoration, the data is further normalized across image channels using a standard deviation of σ=[0.229, 0.224, 0.225] and a mean of μ=[0.485, 0.456, 0.406]. Besides the horizontal flip applied during training, no additional image augmentations are performed to match the input requirements of EfficientNet.

B. Network Details

Fig. 1 shows our joint network architecture, with image transformations detailed between components. The overall functionality is as follows:

- **Data Preparation**: Compressed images are batched and processed according to the steps outlined in the "Data Preparation" section.
- **Image Restoration**: The processed images are input into SwinIR as tensors, which outputs the restored images. The Charbonnier loss term L_A is computed by comparing these restored images with the original ground truth images.
- **Normalization**: After restoration, the images are normalized using standard deviation and mean values before being fed into EfficientNet. EfficientNet then produces a normalized tensor representing the likelihood of the input image belonging to a specific class within the ImageNet-1k dataset. This tensor is compared to the ground truth label to generate the cross-entropy loss term L_B.
- **Loss Aggregation and Model Update**: A linear combination of the Charbonnier loss L_A and the cross-entropy loss L_B is used to update the weights of the SwinIR model. The total loss is the sum of both terms, with a constant ω determining their relative contribution to network backpropagation. To ensure an optimal balance between the two losses, we conducted a series of experiments by sampling network outputs with varying ω values. The final choice of $\omega = 0.99$ reflects a deliberate trade-off, ensuring that the network effectively learns features beneficial for both image restoration and classification tasks.
- **Classification Model**: For the classification network, we selected EfficientNet-B0 due to its compact size and minimal data preprocessing requirements, which involves standard image crops. This choice is particularly relevant given that our joint network includes SwinIR, a computationally expensive model. However, it is important

TABLE I. IMAGENET-1K CLASS MAPPINGS.

Class Label	Class Name
n01440764	tench
n01530575	brambling
n01601694	dipper
n01641577	bullfrog
n01682714	chamelon
n01698640	american alligator
n01740131	night snake
n01770081	daddy longlegs
n01795545	black grouse
n01820546	lorikeet

to note that the specific choice of classification model is not critical to our method. During training, the classification module is frozen, and EfficientNet-B0 serves solely as a validation tool for our approach. While other lightweight models like MobileNet or SqueezeNet could also have been used, EfficientNet-B0 was chosen for its balance between simplicity and efficiency in this context.

III. PERFORMANCE EVALUATION

In this section, we compare the proposed joint restoration-classification network with the existing image restoration methods. All methods evaluate their PSNR, SSIM scores and classification accuracy.

A. Dataset

The training dataset employed in our work was generated from ImageNet-1k [16], which is a subset of a large image database organized according to the WordNet hierarchy. The images in this dataset are quality-controlled and hand-labeled. It is commonly used for training deep learning models for computer vision tasks. ImageNet-1k spans 1,000 object classes and contains 1,281,167 training images, 50,000 validation images, and 100,000 test images. For our research, we randomly selected 10 classes from ImageNet-1k to expedite our experiments. The classes chosen are shown in Table I and Fig. 2.

B. Experimental Setups

The experiments were conducted using a well-defined training configuration. Input images were resized to [3, 224, 224], corresponding to three color channels and a spatial resolution of 224x224 pixels. The model was trained for 70,000 iterations, approximately equivalent to 87 epochs, with a batch size of 16, distributed across 16 nodes with 1 image per node. For data preprocessing, 64 CPUs were utilized, each equipped with 4 GB of RAM, while model training was handled by 16 GPUs, each with 16 GB of VRAM. The Adam optimizer was used. The loss function was a combination of Charbonnier loss and cross-entropy, balanced by a constant ω of 0.99. An initial learning rate of 2e-4 was employed, and a multi-step learning rate scheduler was applied. No additional regularization techniques were applied during the training process. Our dataset was compressed at four different JPEG compression levels (QP = 10, 20, 30, 40), and for each resulting dataset, separate model was trained.

C. Experimental Results

We evaluated the performance of our network across different image qualities using our test set, consisting of 100 images, with consistent image processing steps except for the exclusion of random horizontal flipping. Table I presents the results of the performance metrics, comparing the original compressed images with the newly restored images.

As shown in Table II, images compressed at different levels achieved lower accuracy compared to the original uncompressed images. Another key observation is that the images restored by our approach exhibited a substantial improvement in top-1 accuracy compared to the compressed images. PSNR and SSIM values were consistently higher for restored images across both RGB and YCbCr channels compared to compressed images. Despite lower PSNR and

Fig. 2. Example images of the selected classes from ImageNet-1k.

TABLE II. NETWORK TEST RESULTS COMPARISON.

Image Type	JPEG Compression Quality	Accuracy (%)		PSNR (dB)		SSIM	
		Top-1	*Top-5*	*RGB*	*YCbCr*	*RGB*	*YCbCr*
Original Images	~	82.8	97.2	-	-	-	-
Compresed Images	Level = 10	72.4	89.8	25.00	27.85	0.708	0.778
	Level = 20	77.0	94.6	26.93	29.84	0.788	0.849
	Level = 30	78.80	96.0	28.04	31.08	0.825	0.881
	Level = 40	79.8	95.6	28.84	32.02	0.847	0.900
Our Restored Images	Level = 10	84.6	94.8	27.61	30.51	0.815	0.861
	Level = 20	85.8	96.4	29.85	32.92	0.878	0.916
	Level = 30	85.0	95.8	31.14	34.39	0.904	0.937
	Level = 40	**86.0**	**96.0**	**32.01**	**35.45**	**0.918**	**0.949**

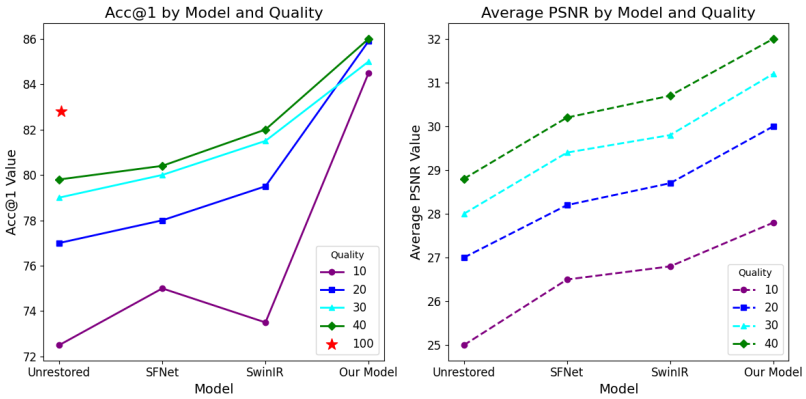

Fig. 3. Comparison of (Left) top-1 accuracy and (Right) average PSNR for different quality levels using our model and the baseline results.

SSIM at reduced image quality levels, the classification accuracy of images restored by our method remained high, ranging from 84.6% at JPEG compression quality level 10 to 86.0% at level 40. This indicates that classification accuracy is not solely dependent on high PSNR or SSIM. This suggests that our network focuses more on enhancing key image features critical for classification rather than just reducing compression noise.

To demonstrate the efficiency of our method, we compared it against the pretrained SFNet and SwinIR models. As shown in the left of Fig. 3, across all JPEG compression quality levels (10, 20, 30, 40), our model consistently surpasses the pretrained SFNet and SwinIR models in terms of PSNR, demonstrating superior restoration quality that closely approximates the original high-quality images. More importantly, the right side of Fig. 3 highlights that our model not only outperforms the unrestored compressed images but also the images restored by the pre-trained SFNet and SwinIR models at each evaluated quality level. Pretrained versions of SwinIR without the classification component showed lower accuracy, likely due to being trained on a different dataset or the absence of cross-entropy loss feedback. Once again, our results revealed that high PSNR and SSIM do not necessarily correspond to high classification accuracy. Training SwinIR jointly with EfficientNet shifted the network's focus towards producing images that maximize classification accuracy, even with a loss function that heavily weighted Charbonnier loss (99%) over cross-entropy loss (1%).

As it can be seen in Fig. 3, our model's top-1 accuracy remains remarkably stable across different image quality levels,

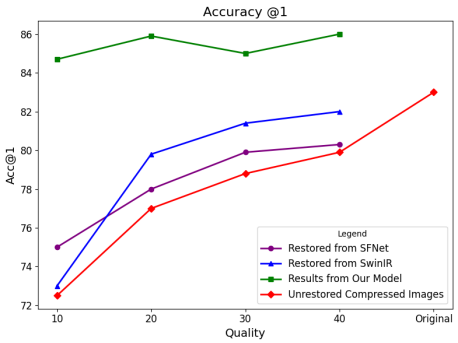

Fig. 4. Comparison of top-1 accuracy metrics for baseline and our results across different image levels.

with a narrow variation of just 1.4% (from 84.6% to 86%). We also observed that our model achieves higher classification accuracy at all four quality levels compared to the original uncompressed images (82.8%), showcasing its exceptional capability for classification tasks. This improvement may be attributed to our joint model's ability to enhance key image features that are critical for classification. Typically, higher image quality leads to better classification accuracy. However, this is not always true. As shown in Fig. 3, images restored by our model achieve higher classification accuracy than even uncompressed originals. This demonstrates that our method enhances features crucial for classification, prioritizing task-specific improvements over general visual quality.

Fig. 4 provides a better comparison of the robustness of the three methods by illustrating their accuracy across different image quality levels. As it can be seen, our model consistently achieves the highest accuracy across all quality levels, outperforming the pretrained SFNet and SwinIR, demonstrating greater robustness.

IV. Conclusion

In this paper, we introduced a novel approach for enhancing compressed images specifically for machine consumption through a joint restoration-classification network. Traditional image enhancement methods often prioritize human visual quality, which does not necessarily align with the needs of machine learning models. Our proposed network addresses this gap by integrating an image enhancement module with a classification module, optimizing both the restoration of image quality and the performance of machine learning tasks. The combination of Charbonnier and cross-entropy loss functions allows our model to balance high-fidelity image restoration with improved classification accuracy. Our experiments demonstrated that this approach effectively mitigates the performance degradation caused by image compression. On the ImageNet-1k dataset, our method achieved substantial improvements in top-1 classification accuracy, with increases of 6.2% for JPEG compressed images at quality level 40 and 12.2% for images at quality level 10, compared to the baseline performance on the same compressed images without enhancement. These results underscore the effectiveness and robustness of our joint network in enhancing compressed images for deep learning applications, highlighting its potential to contribute to more robust and accurate machine vision systems.

REFERENCES

[1] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo and W. Ren, "Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding," in *IEEE Transactions on Image Processing*, vol. 30, pp. 4985-5000, May 2021.

[2] X. Guo, Y. Li and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," in *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982-993, Feb. 2017.

[3] Y. Wang, W. Song, G. Fortino, L. -Z. Qi, W. Zhang and A. Liotta, "An Experimental-Based Review of Image Enhancement and Image Restoration Methods for Underwater Imaging," in *IEEE Access*, vol. 7, pp. 140233-140251, July 2019.

[4] Z. Zou, K. Chen, Z. Shi, Y. Guo and J. Ye, "Object Detection in 20 Years: A Survey," in *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257-276, March 2023.

[5] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," in *Neural Computation*, vol. 29, no. 9, pp. 2352-2449, Sept. 2017.

[6] K. Muhammad, A. Ullah, J. Lloret, J. D. Ser and V. H. C. de Albuquerque, "Deep Learning for Safe Autonomous Driving: Current Challenges and Future Directions," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4316-4336, July 2021.

[7] Y. Zhu and S. Newsam, "DenseNet for dense flow," in *IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 2017, pp. 790-794.

[8] F. He, T. Liu and D. Tao, "Why ResNet Works? Residuals Generalize," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5349-5362, Dec. 2020.

[9] M. M. Kabir, A. Q. Ohi, M. S. Rahman and M. F. Mridha, "An Evolution of CNN Object Classifiers on Low-Resolution Images," in *IEEE International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET)*, Charlotte, NC, USA, 2020, pp. 209-213.

[10] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, 2016, pp. 1-6.

[11] J. Stock, A. Dolan, and T. Cavey, " Strategies for robust image classification, " 2020, arXiv:2004.03452. [Online]. Available: https://arxiv.org/abs/2004.03452

[12] D. Liu, B. Cheng, Z. Wang, H. Zhang, and T. S. Huang, "Enhance visual recognition under adverse conditions via deep networks," in *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4401-4412, Sept. 2019.

[13] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, " SwinIR: Image Restoration Using Swin Transformer," in *IEEE Conference on Computer Vision Workshops (ICCVW)*, Mentreal, Canada, 2021, pp. 1833-1844.

[14] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *International Conference on Machine Learning (ICML)*, Long Beach, California, 2019, pp. 6105-6114.

[15] Y. Cui et al., "Selective frequency network for image restoration," in *International Conference on Learning Representations*, 2023.

[16] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248-255.