HIGH SCHOOL EDITION
Journal of Student Research

Volume 12 Issue 4 (2023)

Image Quality Enhancement via Machine Learning: A Unified Approach to Super-Resolution, Denoising, and Low-Density Enhancement

Woochan Jung[1] and John Blofeld-Watson[1#]

[1]St. Mary's International School – Tokyo, Japan
[#]Advisor

ABSTRACT

Problem: The escalating demand for high-quality images across various applications has underscored the necessity for advanced image enhancement techniques. Traditionally, denoising, super-resolution, and low-density enhancement, the three key image enhancement techniques, have been approached independently, resulting in separate developments for each. Unfortunately, a unified framework that seamlessly combines all three techniques and surpasses individual method performance has been lacking.

Proposed Idea: The objective of this research is to develop a unified image enhancement framework that not only unites these techniques but also substantiates its superiority over existing individual methods through an extensive series of experiments. The proposed method utilizes cascade autoencoder architectures to generate high-quality enhanced images. In addition, an auxiliary artifact type prediction module has been introduced to enhance the noise-awareness, resulting in improved accuracy.

Result: The proposed method demonstrates superior performance in achieving state-of-the-art accuracy when evaluated against three image quality metrics on various public datasets. Additionally, the practical application of the proposed method showcases its efficacy in effectively solving real-world problems.

1. Introduction

1.1 Image Enhancement

Image Enhancement is the process of improving the quality and visual appearances of an image by applying techniques that transform the image. The ultimate goal of this technique is to modify the quality and appearances of an image in a specific way that makes the image more visually appealing. Image enhancement can be applied to various kinds of situations, such as images taken underwater, for the use of the military, and also for driverless car techniques.

1.2.1 Super Resolution

Super-resolution is a fundamental concept in the field of image processing that focuses on enhancing the level of detail and quality in images. The goal of super-resolution techniques is to generate high-resolution images from one or more low-resolution source images. In essence, it aims to improve the visual clarity and sharpness of images, making them more detailed and visually appealing.

Traditional super-resolution methods are based on statistical or optimization-based techniques. These methods can be categorized into two types: interpolation-based and reconstruction-based. Interpolation-based methods simply interpolate the missing high-frequency details using nearby pixels in the low-resolution image. On the other hand, reconstruction-based methods recover high-frequency details by solving an optimization problem that minimizes the difference between the low-resolution image and the high-resolution image.

In order to resolve the aforementioned problems, there have been a number of attempts before our research that proposed promising methods. To start off, super resolution research done by Kim et al. (Kim et al. 2016) proposed a super resolution technique that implemented a very deep convolutional network inspired by VGG-net (Simonyan et al. 2014). Shi et al. proposed a video super-resolution method based on sub-pixel convolutional neural networks (Shi et al. 2016). Their work achieved comparable performance on both still images and videos.

1.2.2 Image Denoising

The goal of image denoising is to remove noise from a noisy image while preserving the underlying signal. Over the years, many approaches have been proposed to address this problem, including both classical and machine learning-based methods. Image denoising is considered as a technique that has an equivalent significance with super resolution.

Lefkimmiatis et al. proposed a novel network architecture for learning discriminative image models that are employed to efficiently tackle the problem of grayscale and color image denoising (Lefkimmiatis et al. 2018). Their method shows comparable results to those of current state-of-the-art networks, despite using a more shallow architecture with significantly fewer trained parameters. Similarly, Gu et al. proposed a self-guided network, which adopts a top-down self-guidance architecture to better exploit image multi-scale information (Gu et al 2019). This method directly generates multi-resolution input images with the shuffling operation in order to extract large-scale contextual information. Finally, Yue et al. explored a different approach to blind image denoising, proposing a new variational inference method that integrated both noise estimation and image denoising into a unique Bayesian framework (Yue et al. 2019).

1.2.3 Low-Density Enhancement

The aim of Low-density enhancement is to recover missing or corrupted information in images that have low-density pixels. One of the previous methods is a study conducted by Guo et al. In this study they constructed the Curvenet, which estimates a non-linear transformation curve of the input image using a zero-reference approach (Guo et al. 2020). The experimental results show that their work outperforms several state-of-the-art methods in terms of both subjective and objective evaluation metrics. Jiang et al. proposed an unsupervised generative adversarial network that can be trained without low/normal-light image pairs (Jiang et al. 2021). Instead of training the network in a supervising approach, their method regularizes the unpaired training using the information extracted from the input itself.

The individual studies conducted in various aspects of image enhancement offer valuable insights within their respective domains, yet a holistic exploration encompassing all three facets – super resolution, image denoising, and low-density enhancement – has been notably absent. Despite the notable advancements achieved by singular studies in image enhancement, a comprehensive investigation that simultaneously addresses all three remains absent.

1.3 Proposed Approach

To the best of our knowledge, a unified framework that can provide all three different types of image enhancement techniques - super-resolution, image denoising, and low-density enhancement - has not yet been proposed. In this paper, I propose a unified image enhancement framework for image denoising, low density enhancement and super-resolution simultaneously in a single pass. The framework is composed of two main parts: the low density and noise artifact resolving part and the super resolution part. Firstly, the low density and noise artifact resolving part determines whether the input image has a noise or a low density. The input image is first fed to the encoder, and outputs the latent code. Then, the type of artifact is identified. After the type of artifact is identified, the latent code goes into the decoder, and the decoder generates an image that does not include noise or low density features of the image. The second part is the super resolution part, which turns the low resolution image to a high resolution image. The architecture in the second part is also similarly done with the first part of the architecture. For the encoder part of the architecture, Similarly with the first part, the image goes into the encoder of the super resolution part of the artifact, and the latent variable is made. The latent variable is then fed to the decoder of the super resolution part, and the final image is returned.

2. Proposed Unified Image Enhancement Framework

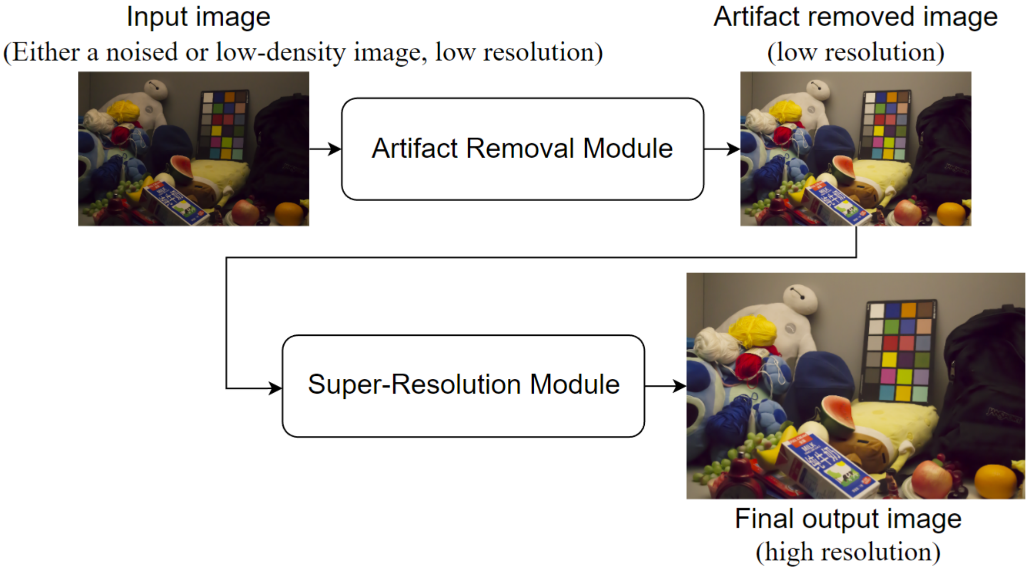

Figure 1. The overall architecture of the proposed unified image enhancement framework

Figure 1 shows the overall architecture of the proposed unified image enhancement framework. As shown in the diagram, the proposed system consists mainly of artifact removal modules and a super resolution module. The primary objective of the artifact removal module aims to eradicate noise and low-density present in the input image. The artifact-removed image is subsequently fed into the super-resolution module, which enhances the resolution of the image and outputs a higher resolution image as the final output.

2.1 Artifact Removal Module

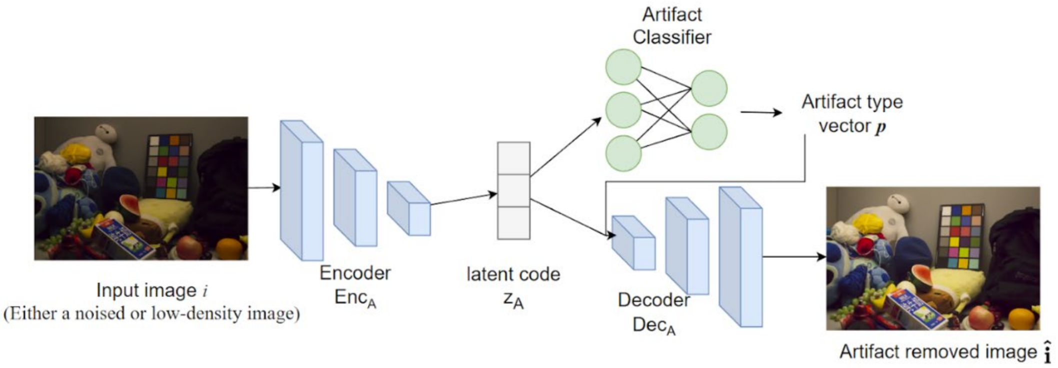

Figure 2. The architecture of Artifact Removal Module

Figure 2 demonstrates the overview of the architecture of the proposed artifact remove module. The proposed artifact removal module consists of encoder, artifact classifier, and decoder. Firstly, the encoder receives the input image $\boldsymbol{i}$ as input and generates latent code z_A. The latent code is then fed to both artifact classifier and decoder for different purposes. The artifact classifier produces artifact type vector $\boldsymbol{p}$ which is input to the decoder. The decoder ultimately produces artifact removed image **i**.

The decoder in the proposed artifact removal module is capable of performing either low-density enhancement or denoising. Hence, the specific action to be performed is determined by the vector $\boldsymbol{p}$ which predicts the artifact type contained in the input image $\boldsymbol{i}$.

The probability vector, denoted as $\boldsymbol{p}$, determines the decoder's prediction regarding the characteristics of the input image. For instance, when $\boldsymbol{p}$={0.8,0.2}, it indicates that the input $\boldsymbol{i}$ is predicted to have a low-density artifact. Alternatively, when $\boldsymbol{p}$={0.1, 0.9}, it suggests that the input $\boldsymbol{i}$ is predicted to contain noise.

By utilizing the proposed network architecture, the decoder can effectively discern the type of artifact contained in the input image, which enables it to generate more natural noise removed images. Experimental evidence to substantiate this claim will be presented in Chapter 4.

2.2 Super-Resolution Module

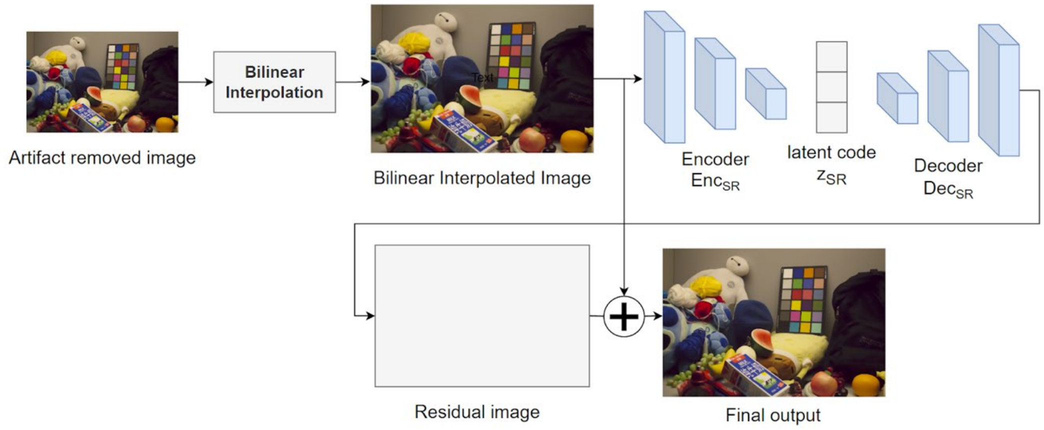

Figure 3. The architecture of super-resolution module

Figure 3 shows the architecture of the proposed super-resolution module. The proposed super resolution module consists of bilinear interpolation, ENC_{SR} (Encoder), and DEC_{SR} (Decoder), The **i** (Artifact removed image) is input to the super-resolution module. Through the process of bilinear interpolation, the pixel values of the image are modified, resulting in I_B (Bilinear interpolation image) that appears slightly blurry compared to the original image. Note that $\boldsymbol{i} \in \boldsymbol{R}^{HxW}$ and $I_B \in \boldsymbol{R}^{2Hx2W}$, where $\boldsymbol{H}$ and $\boldsymbol{W}$ are the height and width of the input image, respectively. The I_B is then fed into the ENC_{SR}.

The ENC_{SR} takes the I_B as input and produces a Z_{SR} as output. This Z_{SR} is then passed as input to the DEC_{SR}. The DEC_{SR} generates I_R as output. The output is defined as $I_R \in \boldsymbol{R}^{2Hx2W}$, where $\boldsymbol{H}$ and $\boldsymbol{W}$ are the height and width of the input image, respectively.

This I_R is added to I_B, combining their pixel values. As a result, $\hat{I}$ (Final output) is produced with higher resolution. $\hat{I}$ is calculated as following:

Equation 1: Final image output calculation

$$\hat{I} = I_R + I_B$$

Throughout the proposed model, two loss functions were mainly implemented to train the system: L_1 loss and L_{ce} (Cross Entropy Loss). After the original image goes through the artifact removal model, the L_1 loss function measures the difference between the original image and the I(the image after the artifact removal model). While the L_1 Loss calculates the difference between the images, the L_{ce} (Cross Entropy Loss) calculates if the predicted noise type vector is correct. Then, 0.9 is multiplied to the value given out by L_{ce} and then added to the value of L_1, giving out the final loss.

Equation 2: L1 Loss Function

$$L_1 = \frac{1}{XY} \sum_{x}^{X} \sum_{y}^{Y} \left| I(x, y) - \hat{I}(x, y) \right|$$

In the equation, the pixels of *I*, which is the image before going into the Artifact Removal Model, is compared to the pixels of I , which is the image generated after the Artifact Removal Model. The difference between the pixel values are measured, and the differences throughout the image are all added up, before getting divided by the number of pixels. The final value indicates how similar or different the two images are before and after going through the Artifact Removal model.

Equation 3: Cross Entropy Loss Function

$$L_{ce} = -\log_e p$$

After the input image is input to the encoder of the Artifact Removal Model, the latent code is generated. The latent code is then fed to the artifact classifier, which gives out ***p***, the predicted noise type vector. The ***p*** is then fed into the negative log function, and the final loss value of Cross Entropy Loss is calculated.

Equation 4: Final Loss

$$L = \alpha L_{ce} + L_1$$

The final loss is determined by multiplying to the Cross Entropy Loss value and adding that to the value of L1 Loss function. The value is set to 0.9 in the proposed model, as the proposed model was best functioning when it was set to 0.9.

3. Experimental Results

3.1 Dataset

3.1.1 DIV2K

The DIVerse 2K resolution high-quality images (DIV2K) dataset (Agustsson et al. 2017) is a popular benchmark for image super-resolution tasks, a technique used to enhance the resolution of an image. This dataset includes 1000 images of diverse contents with 2K resolution. The dataset is divided into 800 images for training, 100 for validation, and 100 for testing. It provides low-resolution counterparts of these images, which are artificially downsampled using different degradation models (such as bicubic interpolation, unknown downsampling, or realistic degradation). This allows models to be trained to 'learn' how to upsample a low-resolution image to its high-resolution counterpart. DIV2K also includes corresponding images with different levels of JPEG compression, providing a dataset for models to learn how to remove compression artifacts.

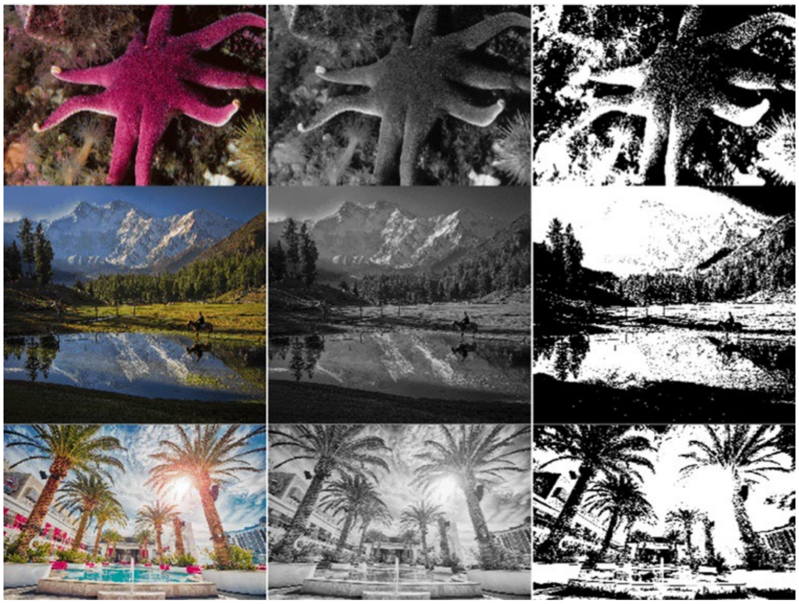

Figure 4. DIV2K examples and their corresponding gray and binary images

3.1.2 DND Dataset

The DND dataset (Plotz et al. 2017), also known as the Darmstadt Noise Dataset, is a widely used benchmark for image denoising. Created by researchers from the Technical University of Darmstadt, Germany, it consists of 50 pairs of noisy and clean images captured under low-light conditions. The dataset contains diverse noise types and levels, including Gaussian, Poisson, and mixed noise, making it challenging for denoising algorithms. Ground truth clean images are provided for evaluation purposes. The DND dataset has played a significant role in the development and comparison of denoising algorithms, serving as a standardized platform for assessing their effectiveness in real-world scenarios.

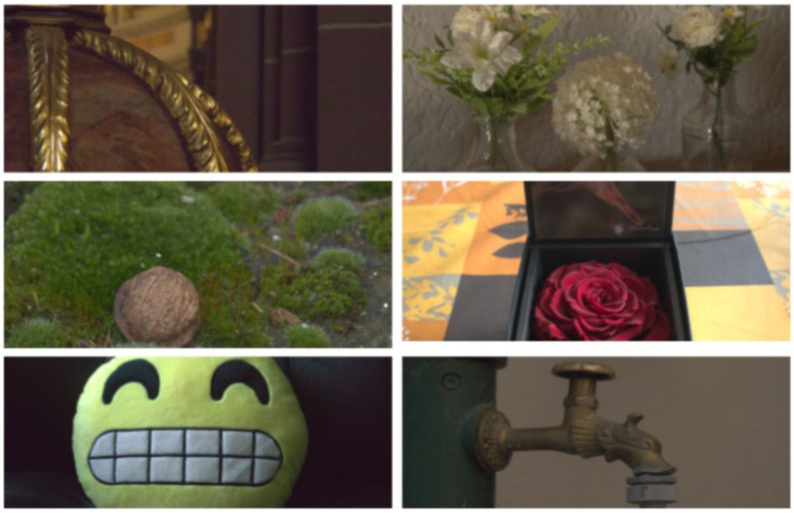

Figure 5. Example samples of DND dataset

3.1.2 ExDark Dataset

The ExDark dataset (Loh et al. 2019) stands as a comprehensive collection of 7,363 images that epitomize low-light conditions. Within this dataset, 12 distinct object classes are meticulously represented: Bicycle, Boat, Bottle, Bus, Car, Cat, Chair, Cup, Dog, Motorbike, People, and Table. Worth noting is the dataset's exclusive focus on images captured under visible light in dim settings.

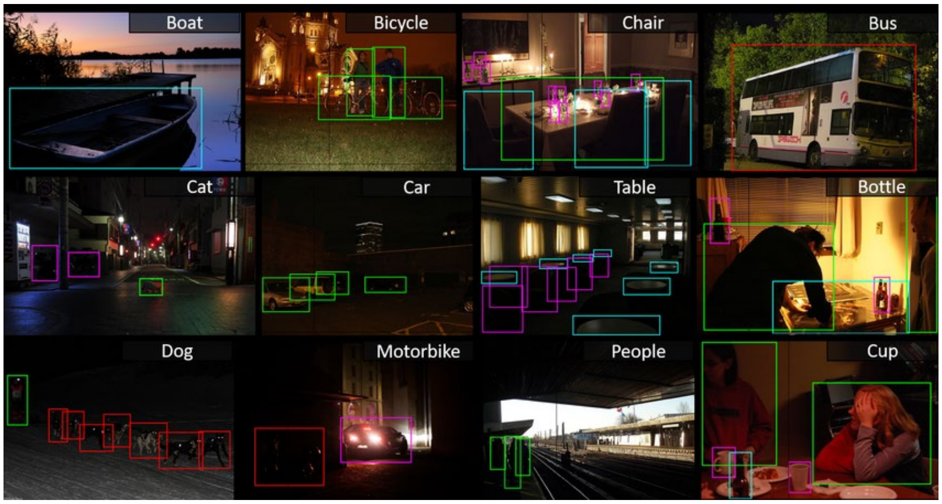

Figure 6. Example samples of ExDark dataset

3.2 Evaluation Metric

3.2.1 Mean Squared Error

Mean Squared Error function is a measure of the average squared difference between the estimated values and the actual values. The goal is to use the equation and determine the loss, and change the model so it has a lower loss. The lower the value of MSE is, the better the model is at predicting the actual object. The ideal value of MSE is 0, which means that the expected value is exactly the same as the actual value.

Equation 5: Mean Square Error

$$MSE = \frac{\sum_{M,N} [I1(m,n) - I2(m,n)]2}{M*N}$$

Where, *M* and *N* refers to the height and width of the image. I_1 is the original low-light image and I_2 is the reconstructed normal-light image. The equation is ultimately calculating the average pixel difference squared.

3.2.2 PSNR

The PSNR (Peak Signal-to-Noise Ratio) is used as an image quality measurement between two images, the original image and the reconstructed image. PSNR aims to reflect on human cognitive and visual perspectives. Equation 5 is an equation for PSNR.

Equation 6: PSNR

$$PSNR = 10\, log_{10}(\frac{R^2}{MSE})$$

Where, MSE refers to the *MSE* value and *R* refers to the maximum pixel value from the images. The higher the PSNR, the reconstructed image is more similar to the original image.

3.2.3 SSIM

SSIM (Wang et al. 2004) stands for Structural Similarity Index Measure, an evaluation metric used to measure the similarity between two images. SSIM takes into account the structural information of the image, such as edges and textures, as well as the brightness and exposure of light. SSIM computes a score between 1 and 0, where 1 shows that the two images are completely identical, and 0 shows that the two are completely dissimilar. Equation 7 is an equation for SSIM.

Equation 7: SSIM

$$SSIM(x, y) = \frac{\left(2\mu_x \mu_y + c_1\right)\left(2\sigma_{xy} + c_2\right)}{\left(\mu_x^2 + \mu_x^2 + c_1\right)\left(\sigma_x + \sigma_y + c_2\right)}$$

The SSIM metric is defined by comparing two images, wish 'x' representing the original image and 'y' representing the reconstructed image. The SSIM function takes into account various aspects of image comparison, including the mean(μ) and variance(σ^2) of both images, as well as their covariance(σxy). Additionally, small constants like c1 and c2 are incorporated to prevent divisions by 0.

3.3 Comparison

To demonstrate the efficacy of the proposed technique, I conducted extensive comparative analysis against state-of-the-art methods that tackle each super-resolution, denoising, and low-density enhancement task separately.

Table 1. Comparison with state-of-the-art super-resolution methods (DVI2K)

Method	**PSNR**	**SSIM**
SRCNN (Dong et al. 2015)	33.05	0.9581
VDSR (Kim et al. 2016)	33.66	0.9625
EDSR (Lim et al. 2017)	35.12	0.9699
Ours	36.08	0.9895

Table 1 shows how the state-of-the-art super-resolution methods show different performances on each evaluation metric. Across all evaluation metrics—PSNR and SSIM—the proposed method consistently outperforms the other two state-of-the-art methods, clearly demonstrating its superior ability to convert the original image into a high-resolution version.

Table 2. Comparison with state-of-the-art denoising methods (DND)

Method	PSNR	SSIM
CBDNet (Guo et al. 2019)	38.06	0.942
RIDNet (Anwar et al. 2019)	39.26	0.953
AINDNet (Kim et al. 2020)	39.37	0.951
Ours	39.35	0.952

Denoising performance was tested using DND dataset. We compared the denoising capabilities of the proposed method to the existing state-of-the-art denoising methods. The proposed method shows comparable performance in terms of PSNR and SSIM.

Table 3. Comparison with state-of-the-art low-density enhancement methods (ExDark)

Method	PSNR	SSIM
DSLR (Lim et al. 2020)	15.05	0.597
Retinex-Net (Wei et al. 2018)	16.77	0.462
LLNet (Lore et al. 2017)	17.95	0.713
Ours	18.01	0.778

For low-density enhancement, ExDark dataset is used for comparison. For the comparison methods, I choose DSLR (Lim et al. 2020), Retinex-Net (Wei et al. 2018), and LLNet (Lore et al. 2017) which show comparable performance in low-density enhancement tasks. As shown in Table 3, the proposed method shows notable proficiency in enhancing low-density areas of images, surpassing the performance of the existing state-of-the-art methods.

Table 4. Comparison with state-of-the-art methods (Unified Dataset)

Method	PSNR	SSIM
AINDNet (Kim et al. 2020) + LLNet (Lore et al. 2017) + EDSR (Lim et al. 2017)	19.84	0.697
Ours	25.95	0.810

Finally, an additional experiment is conducted to assess the efficacy of the proposed method in addressing the challenge of unified image enhancement. Initially, I present two distinct problems. The first problem involves the fusion of super-resolution and denoising tasks, while the second problem encompasses low-density enhancement and super-resolution.

For the initial task, the input comprises a low-resolution image with added noise, while the desired output is a high-resolution image with the noise effectively removed. In the case of the second task, the input involves a low-resolution image with low density, and the goal is to generate a high-resolution image with improved density.
To facilitate a comprehensive evaluation, three state-of-the-art methods are selected—each excelling in a specific task—and these are combined for comparative analysis. As depicted in Table 4, the proposed methods demonstrate a substantial performance superiority over the amalgamation of state-of-the-art approaches. This outcome unequivocally underscores the enhanced effectiveness of the proposed unified framework.

3.4 Application

To demonstrate the applicability of the proposed method in real-world scenarios, I conduct additional experiments that aim to validate its performance under various conditions and challenge situations. I consider the proposed image enhancement framework as a valuable preprocessing step for Object Detection, a highly utilized machine learning task with real-world applications like autonomous driving. Initially, I evaluate the performance of a pre-trained object detection method on artifact-laden images, characterized by low density or noise. Subsequently, these artifact-affected images are input into the proposed framework. The resulting artifact-free, high-resolution images are then channeled into the same object detection methods, allowing for a direct comparison of performance enhancements.

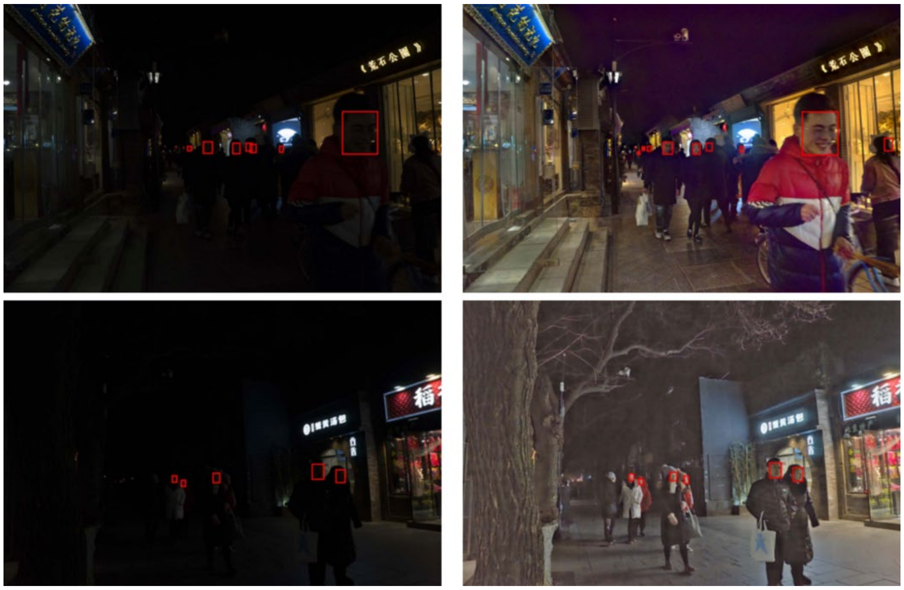

Figure 7. Object detection with proposed image enhancement method

Table 5. Object detection result with proposed image enhancement method

Method	Bounding Box Average Precision	
	Proposed method not applied	Proposed method applied
YOLO (Redmon et al. 2016)	30.5	48.8
SSD (Liu et al. 2016)	31.8	59.7
RetinaNet (Lin et al. 2017)	35.9	67.9

As depicted in Figure 7 and summarized in Table 5, the impact of the proposed image enhancement framework on object detection tasks is evident. The findings strongly suggest that the proposed method has the potential to significantly elevate the accuracy of off-the-shelf object detection models when applied in real-world scenarios. This enhancement holds the promise of fostering advancements across diverse industries, showcasing the versatility and practicality of the proposed approach.

4. Conclusion

In this paper, I proposed a novel image enhancement network that combines all three techniques: super-resolution, denoising, and low-density enhancement. Machine learning and computer vision have advanced image enhancement techniques, but most focus on single tasks like denoising or super-resolution. Our novel architecture addresses these limitations, handling multiple tasks simultaneously and producing significantly improved results by optimizing their interplay. As there has not been an unified framework of all three enhancement methods, the main aim of this project was to propose a unified image enhancement framework, and prove that the proposed method outperforms the existing individual methods through extensive experiments. The experiments conducted for the three enhancement techniques shows that the proposed method has an outstanding performance compared to other models. Thus, it can be concluded that the aim to propose a unified framework of all three techniques has come to a success. I expect the proposed model to be actually implemented in real-life situations, and make a great contribution to a diversity of fields.

References

Agustsson, E., & Timofte, R. (2017). Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 126-135).

Anwar, S., & Barnes, N. (2019). Real image denoising with feature attention. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3155-3164).

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.

Gu, S., Li, Y., Gool, L. V., & Timofte, R. (2019). Self-guided network for fast image denoising. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 2511-2520).

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., & Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1780-1789).

Guo, S., Yan, Z., Zhang, K., Zuo, W., & Zhang, L. (2019). Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1712-1722).

Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., ... & Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing, 30, 2340-2349.

Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

Kim, Y., Soh, J. W., Park, G. Y., & Cho, N. I. (2020). Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3482-3492).

Lefkimmiatis, S. (2018). Universal denoising networks: a novel CNN architecture for image denoising. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3204-3213).

Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).

Lim, S., & Kim, W. (2020). DSLR: Deep stacked Laplacian restorer for low-light image enhancement. IEEE Transactions on Multimedia, 23, 4272-4284.

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

Loh, Y. P., & Chan, C. S. (2019). Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding, 178, 30-42.

Lore, K. G., Akintayo, A., & Sarkar, S. (2017). LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61, 650-662.

Plotz, T., & Roth, S. (2017). Benchmarking denoising algorithms with real photographs. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1586-1595).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560.

Yue, Z., Yong, H., Zhao, Q., Meng, D., & Zhang, L. (2019). Variational denoising network: Toward blind noise modeling and removal. Advances in neural information processing systems, 32.