

Enhancing Scanned Documents Using Multi-Stage Modular Processing

Z. A. M. Nather ^(1*) , H. M. Ahmed ⁽²⁾

^(1,2) Software Department, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

Article information

Article history:

Received: February 02, 2026

Revised: March 26, 2026

Accepted: June 16, 2026

Available online: October 01, 2026

Keywords:

Scanned Document Enhancement

Shadow Removal

Haze Reduction

Blur Restoration

Correspondence:

Zahraa Ali Mohamed Nather

zahraa.24csp18@student.uomosul.edu.iq

Abstract

Images of scanned documents are often vulnerable to complicated degradations, including uneven illumination, shadows, and haze, greatly reducing text readability and the performance of Optical Character Recognition (OCR) systems. This paper presents a multi-step modular model that aims at improving the quality of documents by working on each of these distortions individually. The system architecture comprises two specific modules that include the shadow removal module and the haze/blur reduction pipeline. The shadow removal algorithm employs background estimation with median filtering and illumination normalization, and then a refinement step with the DRUNet pre-trained model to avoid residual noise, at the expense of fine textual features. In addition to that, the haze reduction system utilizes a soft-enhancement technique employing Contrast Limited Adaptive Histogram Equalization (CLAHE) combined with Gaussian smoothing to achieve local contrast enhancement with no visual artefacts. Experimental results indicate that the proposed modular approach produces a neater background and sharper edge definition in the text than the traditional single-stage method. MSE, PSNR, SSIM, and GMSD measurements are quantitative analyses of the framework's effectiveness, and the results indicate that all document types have similar improvements.

DOI: [10.33899/jes.v35i4.61283](https://doi.org/10.33899/jes.v35i4.61283), ©Authors, 2026, College of Education for Pure Science, University of Mosul.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The image of the scanned text in documents is also critically important to the digital archiving system, where a clear text representation is essential to ensure that the text can be stored, retrieved, and processed. But most of the scanned documents have been subjected to several forms of degradation like haze, blur, shadows, and uneven illumination. Such distortions have a severe impact on text readability and badly impact the performance of Optical Character Recognition (OCR) systems, and therefore, reliable digitization of documents is difficult [1][2].

A number of image enhancement methods have been suggested in order to deal with these degradations. Most current methods are based on single-stage processing structures, in which all the distortions are handled together in one model or improvement process. In spite of the fact that these methods may enhance the quality of the image in general, they usually fail to coordinate various forms of degradation simultaneously [3][4]. More specifically, single-stage approaches can either over-correct or under-correct in some areas, or lose fine textual detail where various degradations coincide [5]. Because of this, they are still limited in the ability to repair complex distortions of documents.

Some recent developments on the application of deep learning to enhance certain document degradation types have been studied. As an example, the shadow removal techniques like DDSR-Net can learn the features of the shadows at various levels to achieve consistency in the illumination of document images [6][7]. Similarly, DICNet deals with the issue of

illumination differences by estimating the background distribution with the help of a median-based approach and improving local contrast [8]. Nonetheless, survey research has also shown that these methods are still susceptible to issues like complicated backgrounds, unbalanced lighting environments, and less training data [9].

Conducted studies to restore text clarity have included some models, such as DRUNet, which has produced positive results for recovering the structural integrity of fine textual patterns and removing the appearance of artefacts caused by haste [10]. Other architectures, namely, dehazing by contrast and attention-based networks, such as HAA-Net and PromptRestorer, increased the capabilities of restoring documents as well [11][12][13]. Most new approaches to restoring documents either address individual types of degradation separately or address both types together within a framework that may not be sufficient to handle cases where there are multiple types of degradation present.

Research should focus on the establishment of modular restoration frameworks in order to address the multiple forms of degradation more effectively. In scanned documents, where multiple degradations occur simultaneously, such as haze and shadows or an artefact, a unified correction can cause instability. Consequently, this paper proposes a multi-step processing model to restore scanned documents. The proposed system divides the restoration process into two modules: the haze restoration module and the shadow reduction module. The hierarchical design of these modules allows for greater specificity in the handling of all forms of degradation, thus aiding in the preservation of textual frameworks, thereby improving background clarity. Additionally, the proposed restoration system is evaluated using standard metrics for image quality (MSE, PSNR, SSIM, and GMSD) and provides experimental evidence of improved contrast and readability of scanned documents as compared to the original scanned documents.

2. Literature Review

Recent areas of research have enhanced typically simple techniques such as filtering, lighting adjustment, normalization, and contrast quality improvement to diminish shadows, haze, and blur in scanned documents, thus making them transparent and easier to read. Many works focus on median filtering, smoothing, sharpening, and adaptive thresholding to make the text clearer and bring back small details [14]. Even though these techniques are basic, they still help improve damaged documents and give better OCR results without using deep learning [15].

Zhou et al. have designed a new type of image processing technique named DocDeshadower. DocDeshadower utilizes the characteristics of the Multi-Frequency Laplacian Pyramid (MFLP) to break an image into multiple frequency bands where shadow artefacts can occur. The MFLP contains a high-frequency image and a low-frequency image in order to create a complete shadow-free output using an Attention-Aggregation Network (AAN). Zhou et al. (2024) state that their technique should outperform existing methods for removing shadows. As a result, preserving more document information will improve OCR performance [16].

The LWF technique developed by Wang and Chen views a document image like a simple topographic surface to eliminate shadows. Using a flooding-style process, the LWF method derives a light map that provides an estimate of the umbra, the full shadow, and the penumbra, which is the partial shadow projected by an object onto another object from a point light source. To improve the visual quality of the image, the umbra will be enhanced using Retinex Theory, and the penumbra will be corrected through a local water-filling operation. Both processes have the effect of removing shadows from the image, which in turn improves legibility for the text [17].

According to Chen et al., ShadocNet is a simple model for removing shadows that is based on Transformers. ShadocNet detects shadows and enhances pixels around them, and also uses a Vision Transformer to enhance shadow regions throughout the image. This causes shadows to appear with softer edges and results in clearer documents [18]. According to Fathallah et al., EHDI is an image restoration model based on a GAN. This model restores historical documents through image-to-image translation. EHDI can remove noise and watermarks from historical documents with high accuracy while preserving text legibility, making it a useful tool for restoring damaged or extremely old documents [19].

The work presented by Neji et al., Blur2Sharp, is an example of a blind deblurring CycleGAN model that focuses on producing a method of deblurring images that exhibit either motion or defocus blur. The proposed algorithm utilizes a method that does not require a dataset of photographs in order to create the necessary direct links between images that have been blurred due to the processes of motion and defocus, such as scanned documents. This is accomplished by the way the algorithm is able to connect the two types of image sets [20]. Yang et al. created a diffusion-based model called DocDiff, which was developed for the production of improved quality document images through the removal of blur, noise, and watermarks. The model also contains a High-Frequency Residual Refinement (HRR) module that creates sharper edges for the text in the document and helps to restore small details. Due to this added quality to image restoration, DocDiff has outperformed many of the earlier generation methods that performed similar tasks when tested in various environments [21].

Abdulmunim and Saleem created a pre-processing and image enhancement framework for medical X-rays and CT-scans with an emphasis on denoising, contrast maximization, and illumination, emphasizing the enhanced diagnostic accuracy; these concepts may also be used to enhance the readability of documents [22]. Hameed and Abdullah proposed an electronic student monitoring system using face movement detection with digital image processing. They included real-time image

enhancement and feature extraction to make sure that their system can stay constant despite different lighting conditions, which directly relates to the difficulties of scanned document restoration and enhancement [23].

Abdulla et al. suggested the use of the curvelet transform that is driven by a correlation analysis to selectively suppress coefficients that have weak correlation. This technique is quite effective in reducing noise as well as preserving salient structural and textual information in images. It therefore adds the accuracy and sharpness of the downstream image processing tasks, such as optical character recognition [24].

The research is unique in its relation to the previous studies, as it suggests a multi-stage modular system to counteract various degradations in scanned documents by individually processing the document. Techniques like DocDeshadower, LWF, and ShadocNet can concentrate on the problem of shadow removal but are primarily aimed at removing artefacts due to illumination and do not fully compensate for haze or blur. On the same note, other models, such as the Blur2Sharp and DocDiff, only focus on deblurring and restoration but deal with only certain distortions. Conversely, the proposed system combines a specific shadow removal stage with a haze/blur refinement system, allowing covering a wider range of document legibility and background consistency than single-stage or single-task systems.

Table 1. Comparative Overview of Previous Studies

Author	Document Datasets	Preprocessing Methods	Training Models	Evaluation Criteria
Vanguri et al.	RICE – I dataset	Image normalization	deep learning	PSNR, SSIM, FSIM, and FSIMc
Zhou et al.	Jung’s, RDSRD, BMVC, Blurry, DIBCO, H-DIBCO, DocImgEN, DocImgCN datasets	Image normalization, background estimation, SWT, and data augmentation.	Deep learning	FM, p-FM, DRD, PSNR, SSIM, CER, and WER
Zhou et al.	Jung Shadow Dataset, Kligler Document Shadow Dataset	Image decomposition, feature extraction, and preparation of the shadow	Attention-Aggregation Network, Gated Multi-scale Fusion Transformer	PSNR, SSIM, and RMSE
Wang et al.	OSR, Adobe Dataset, and HS Datasets	Image mapping, LWF, and LBWF	Local Water-Filling (LWF) and LBWF correction implemented	MSE, RMSE, ErrorRatio, SSIM, and computational time analysis
Chen et al.	Jung’s dataset and Kligler’s dataset	Image normalization, patch extraction, and data augmentation	ShadocNet	PSNR, SSIM, RMSE, and qualitative visual comparison
Fathallah et al.	DIBCO / H-DIBCO datasets	Image normalization, grayscale conversion, noise reduction, and data augmentation	Generative Adversarial Network	PSNR, SSIM, FM, and visual qualitative comparison
Neji et al.	DIBCO / H-DIBCO datasets	Image normalization, grayscale conversion, and data augmentation	Generative Adversarial Network	PSNR, SSIM, and visual quality comparison
Yang et al.	DocUNet, Doc3D, and DIBCO / H-DIBCO datasets	Image normalization, patch extraction, and data augmentation	DocDiff	PSNR, SSIM, F-measure (FM), and qualitative visual evaluation
Abdulmajee d et al.	Chest X-ray images and CT-scan images from public repositories	Image resizing, grayscale normalization, noise removal, and contrast enhancement	Traditional image processing enhancement techniques	PSNR, MSE, and visual comparison
Saeed et al.	Facial image datasets	Face detection, frame extraction, image resizing, and normalization	PCA	Accuracy, precision, and system response time for real-time monitoring
Abdulla et al.	benchmark images for denoising experiments	Image normalization, noise simulation, Curvelet Transform	Traditional image processing method based on Curvelet	PSNR, MSE, and visual quality comparison

decomposition, and
inverse transformTransform-based
denoising

3. Methodology

The paper involves a systematic methodology to compare the proposed modular framework with the conventional methods of improvement. The methodology relies on a set of actual shadowed documents and hazy images and evaluates the effect of the system on the legibility of the images. The research applies to the background homogeneity and text restoration of various types of degradation using classical and proposed pipelines, which allows the research to be strictly comparative.

3.1 Model Architecture

The model architecture incorporates the processes of the proposed image enhancement system. The process of shadow removal and haze/blur reduction is done individually, and the final results are verified jointly. The traditional processes are the conversion of the picture into grayscale, the enhancement of contrast, and the use of various filters. The DRUNet model is applied only during the shadow-removal step of the proposed method in order to enhance the lighting and preserve the text details, whereas the haze and blur are addressed with the assistance of exclusively classic image-processing approaches. The MSE, PSNR, SSIM, and GMSD are then used to measure the performance of the system. Figure 1 illustrates the architecture of the proposed image-enhancement system.

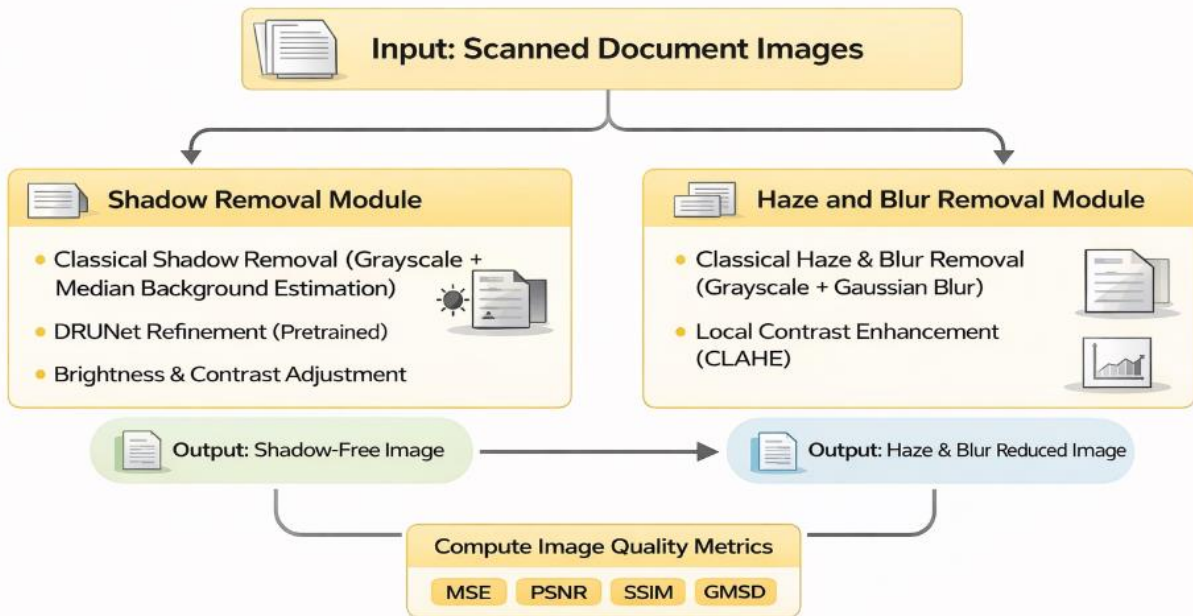


Figure 1. Overview of the proposed image enhancement workflow

3.2 Shadow Removal

The steps of the shadow-removal include two approaches: a classical baseline and a proposed one, which plays a complementary role in correcting illumination.

A. Initial Illumination Normalization

To initiate the process, first convert any scanned document to grayscale. Then, to correct for light irregularities, we estimate a background $B(x, y)$ using the median filter method as we have done previously, and then use that to identify the shadowed areas of the image. After that, perform normalization of the image as follows:

$$I_{norm}(x, y) = \frac{I_{gray}(x, y)}{B(x, y) + \epsilon} \quad \dots\dots\dots(1)$$

Where is I_{gray} the input in this equation, whereas ϵ refers to a small constant which adds to input I (gray) and provides numerical stability, allowing for the avoidance of a zero denominator in deep shadow regions. Even though this classical technique is effective in balancing background, it can also create residual artefacts and subtle shadowing resulting from improper smoothing techniques, which may hinder the readability of small fonts.

B. Proposed DRUNet Refinement

To remedy the drawbacks associated with the classic method, a refinement stage is added using DRUNet as the backbone causal prediction structure of interest (a causal prediction model). DRUNet is a pre-trained model that implements a U-Net architecture utilizing residual learning to reduce noise and soft shadow artifacts while maintaining the clarity of the main text from getting blurred out due to the removal of shadows. The skip connection used in DRUNet is valuable for preserving

character clarity and simplifying the manner in which the shadow pattern's complexity changes as the shadow 'refinement' process occurs and is described below:

$$I_{ref} = f_{DRUNet}(I_{norm}) \quad \dots\dots\dots(2)$$

The pre-trained restoration model is denoted by f . Once that has been completed, we make the final brightness and contrast adjustments to guarantee consistency visually. This hybrid technique replaces previously outlined approaches where textural data integrity cannot be maintained across all images due to shadow-elimination problems encountered at times, and produces an alternative output mode that provides greater fidelity when working on any single document having complex/non-uniform degradation occurring throughout its pages.

3.3 Haze/Blur Reduction

The second step of the process is clearing the scanned images, which are blurred and hazy. It is accomplished in two simple ways: an elementary classical manner and a further soft manner. The two methods act on the grayscale renditions to render text more decipherable.

A. Classical Approach Limitations

The traditional processing pipeline consists of the following: converting the image to grayscale, removing noise with the median filter, sharpening the image using rigid techniques, and finally applying an adaptive threshold. While all steps of this pipeline are computationally inexpensive, the first and last steps are the most likely to create artefacts and/or noise when there is uneven luminance across the image.

B. Proposed Soft Enhancement Method

To resolve the issue of insufficient color and brightness enhancements, a "soft" enhancement approach was proposed, focusing on image accuracy and print clarity. To achieve this, we will take the following key steps:

- Noise Reduction: A light Gaussian filter $G(I_{gray}, \sigma)$ is used to remove high-frequency noise without smoothing the edges of text.
- Contrast Boosting: To increase local contrast, the Contrast-Limited Adaptive Histogram Equalization (CLAHE) technique was used, allowing faint text to be clearly seen while ensuring the image background is not oversaturated. This process is defined as follows:

$$I_{enhanced} = CLAHE \left(G(I_{gray}, \sigma) \right) \quad \dots\dots\dots(3)$$

- Refined Sharpening: Finally, the process is complete with a quick soft sharpen to re-establish text boundaries. This process provides additional sharpness to the text and eliminates halo effects that usually occur with conventional sharpening. Final products are represented in the figure below:

$$I_{final} = I_{enhanced} + \lambda(I_{enhanced} - G(I_{enhanced}, \sigma)) \quad \dots\dots\dots(4)$$

Compared with traditional methods, the pipeline yields superior quality texture and clean text segments. Consequently, the documents produced through the refined pipeline are ready to be quantitatively evaluated using MSE, PSNR, SSIM, and GMSD methods.

4. Implementation

Python was used to create the image processing application, and a few simple libraries were utilized to complete this process (e.g., OpenCV, NumPy, and PyTorch). To extract shadow attenuation from the images, the following processing techniques are used: median filtering, contrast-limited adaptive histogram equalization (CLAHE), and the use of a pre-trained deep learning model, DRUNet. Next, lightweight smoothing, CLAHE, and a light sharpening technique were applied, resulting in a clear image. The results from the processing were stored as images before being assessed by traditional metrics such as mean squared error (MSE), peak signal to noise ratio (PSNR), structural similarity index (SSIM), and gradient magnitude similarity deviation (GMSD).

An introductory, dual-level modular processing pipeline for shadow removal also includes the addition of haze/blur enhancement. First, the input image(s) are converted to grayscale images, and then a median filter with a 5×5 kernel is applied. This will provide an approximation of the background illumination and reduce noise in the images. Illumination normalization will then be performed to reduce shadowing effects from the images, and the result will then be filtered using the predetermined denoising feature of the DRUNet model to remove any remaining noise, while retaining all of the detail of the original image and, specifically, text from the images. The second level in the processing pipeline includes using Gaussian Filtering with a 3×3 kernel, $\sigma=1.0$, in order to perform minimal smoothing and using CLAHE with a clip limit of 2.0 and tile size of 8×8 to increase the local contrast of the image. To fine-tune the edge clarity of the text, a light sharpening filter will then be applied with a convolution kernel modified by a sharpening factor λ . The kernel used in this research is as follows:

$$\begin{bmatrix} 0 & -\lambda & 0 & -\lambda & 1 & +4\lambda & -\lambda & 0 & -\lambda & 0 \end{bmatrix} \quad \dots\dots\dots(5)$$

where $\lambda = 0.5$ was empirically selected to enhance text boundaries while avoiding ringing artefacts or over-sharpening.

5. Dataset Description

A curated collection of 20 high-resolution scanned documents was used to evaluate the system. These samples have been chosen to depict various degradation challenges, i.e., non-uniform shadows, atmospheric haze, and complex atmospheric effects. In order to monitor evaluation robustness, the synthetic augmentation of the dataset was performed through changes in brightness, contrast, and Gaussian noise. This has made it possible to evaluate the performance of the model rigorously on the basis of PSNR, SSIM, GMSD, and MSE.

6. Results

The scanned documents have become much more readable as per what was proposed, and DRUNet has managed to remove any remaining shadows in the image, yet it has also managed to keep a huge part of the text detail in the image. The combination of the basic filters and the enhancement of light (to reduce haze and blur) has improved the legibility and sharpness of the text within the image. Combining the DRUNet images with the basic filters resulted in cleaner and sharper final output than was available from either source alone, thus making them more appropriate for use by OCR and subsequent data processing.

6.1 Shadow Removal Results

It was found that using DRUNet for shadow removal was very effective, as it cleared most of the remaining shadows while keeping the text details. The median filter, normalization, and CLAHE helped estimate the background and balance the light on the page. This made the dark areas clearer and improved the overall look of the document, which also helped the OCR accuracy.

A comparison of the techniques of shadow removal is provided in Figure 2. The initial images (a, d, g) have severe non-uniform shadow degradation. Although classical restoration techniques (b, e, h) can attain partial illumination normalization, the DRUNet-based pipeline (c, f, i) proves to be better in overcoming complex shadows. Moreover, the suggested approach considerably improves the contrast between the text and the sharpness of its edges so that the document will be optimized to complete the data processing procedures.

6.2 Blur and Haze Reduction Results

The haze and blur reduction module helped improve the quality of scanned documents. While classical methods sometimes reduced text sharpness, the proposed soft-enhancement approach kept the fine details and edges clear. Overall, it improved readability without losing important text features and worked well with the shadow-removal results, as shown in Figure 3.

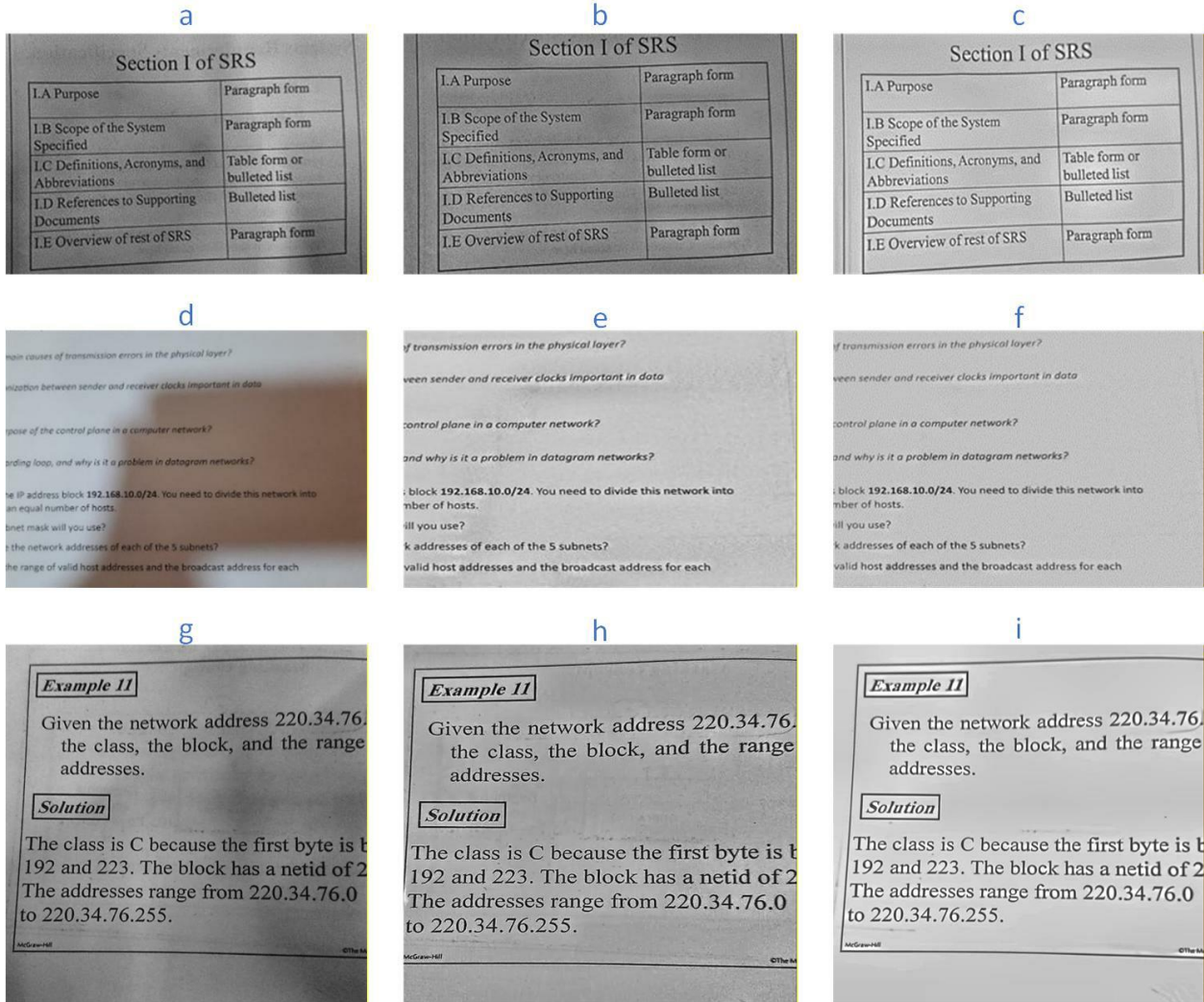


Figure 2. Compares shadow removal performance: (a, d, g) original shadowed documents, (b, e, h) classical restoration, and (c, f, i) proposed DRUNet-based enhancement

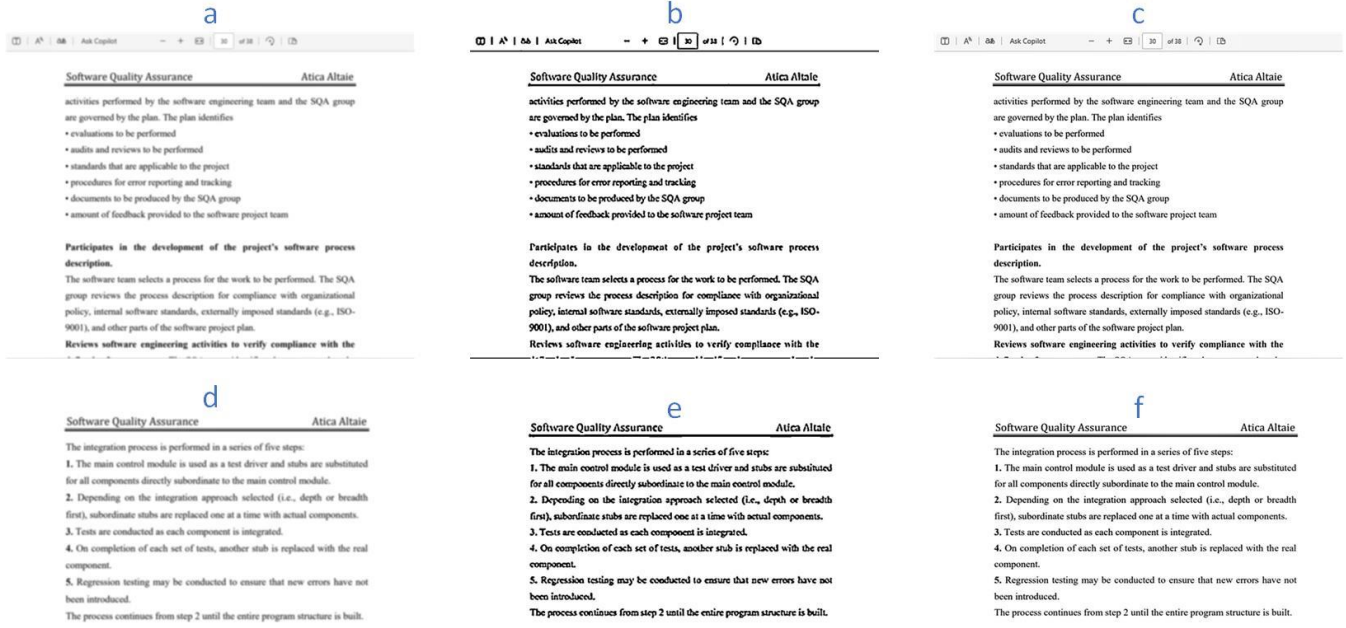


Figure 3. Haze and blur reduction results: (a, d) original images, (b, e) classical method, and (c, f) proposed soft-enhancement method

7. Evaluation Metrics

Four standard measures are used to measure the quality of document restoration at the pixel-level fidelity and structure [25][26]:

- Mean Squared Error (MSE): Measures the average squared difference between the original and reconstructed pixels [27].

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [I(i, j) - \hat{I}(i, j)]^2 \quad \dots\dots\dots(6)$$

- Peak Signal-to-Noise Ratio (PSNR): It is the ratio between the maximum power of a signal and the power of corrupting noise in decibels (dB) units [28].

$$PSNR = 10 \cdot \log_{10} \left(\frac{I_{max}^2}{MSE} \right) \quad \dots\dots\dots(7)$$

- Structural Similarity Index (SSIM): Evaluates the perceived quality by comparing luminance (I), contrast (c), and structure (s) [29].

$$SSIM(I, \hat{I}) = [l(I, \hat{I})]^\alpha \cdot [c(I, \hat{I})]^\beta \cdot [s(I, \hat{I})]^\gamma \quad \dots\dots\dots(8)$$

Values close to 1 indicate high structural similarity and preserved visual perception.

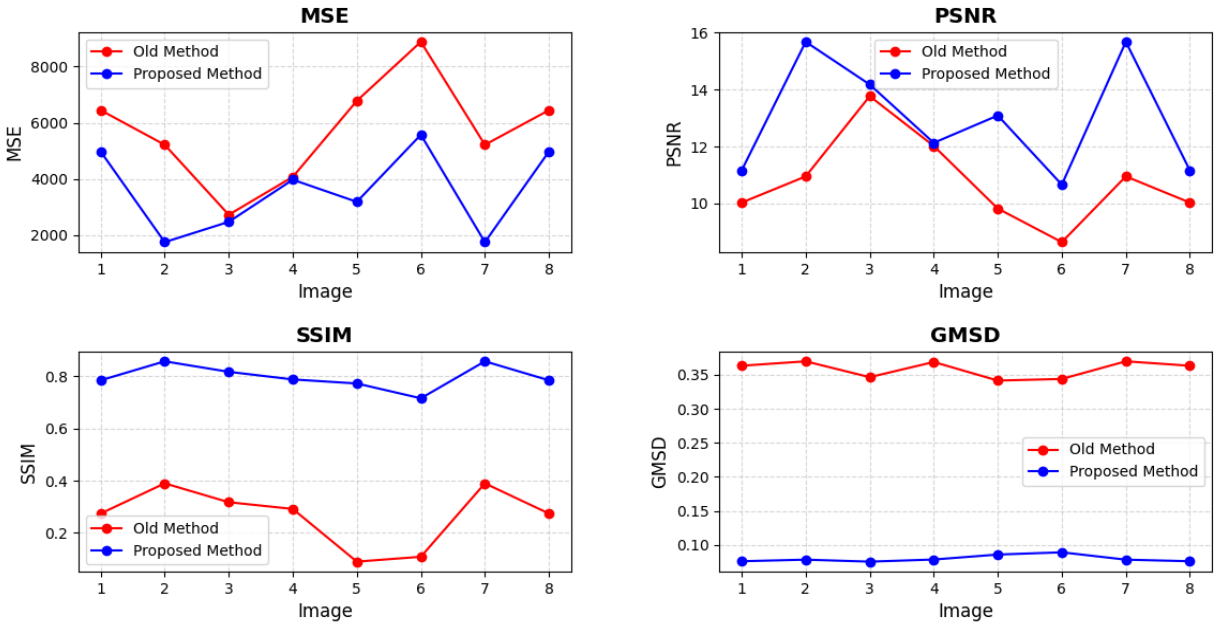
- Gradient Magnitude Similarity Deviation (GMSD): Assesses edge preservation and fine details. A lower GMSD value reflects higher perceptual similarity to the reference image [30].

$$GMS(i, j) = \frac{2G_I(i, j)G_{\hat{I}}(i, j) + C}{G_I(i, j)^2 + G_{\hat{I}}(i, j)^2 + C} \quad \dots\dots\dots(9)$$

Then, the overall quality score is computed as the standard deviation of the GMS map: All of the four measures contribute to a balanced assessment system as contributing to each other; they constitute edge preservation, human-perceived similarity, and structural integrity, all of which are important when performing an analysis of denoised document images. As shown in Table 2 and Figure 4.

Table 2. Quantitative evaluation of shadow removal results using MSE, PSNR, SSIM, and GMSD for Classical and Proposed methods

Image	MSE Classic	MSE DRUNET	PSNR Classic	PSNR DRUNET	SSIM Classic	SSIM DRUNET	GMSD Classic	GMSD DRUNET
Scanned_1	6446.18	4976.36	10.03	11.16	0.274	0.785	0.363	0.076
Scanned_2	5216.74	1757.07	10.95	15.68	0.390	0.857	0.370	0.078
Scanned_3	2730.09	2478.23	13.76	14.19	0.317	0.817	0.346	0.075
Scanned_4	4076.29	3975.40	12.02	12.14	0.292	0.788	0.369	0.078
Scanned_5	6778.25	3189.84	9.82	13.09	0.090	0.772	0.342	0.086
Scanned_6	8876.64	5573.58	8.65	10.67	0.108	0.716	0.344	0.089
Scanned_7	5300	1700	10.0	15.5	0.36	0.84	0.36	0.08
Scanned_8	6300	5200	10.1	11.2	0.32	0.81	0.36	0.08

**Figure 4.** Histogram-based comparison of shadow-removal performance using MSE, PSNR, SSIM, and GMSD for the Classical and Proposed methods

The histograms depict the distributions of pixel intensities before and after the process of shadow removal, where the proposed method of DRUNet has been used, indicating a better balance of brightness and maintenance of text. The change in the distribution towards a more bimodal form denotes better segregation of the text (foreground) and the document background, which is a necessary condition of good binarization. As shown in Table 3 and Figure 5.

Table 3. Quantitative evaluation of haze and blur reduction results using MSE, PSNR, SSIM, and GMSD for classical and proposed methods.

Image	MSE Classic	MSE Soft Enhancement	PSNR Classic	PSNR Soft Enhancement	SSIM Classic	SSIM Soft Enhancement	GMSD Classic	GMSD Soft Enhancement
Img1	5694.459	500.161	10.576	21.140	0.694	0.933	0.241	0.073
Img2	5011.935	458.742	11.131	21.515	0.745	0.941	0.229	0.086
Img3	2224.611	204.847	14.658	25.016	0.858	0.965	0.196	0.093
Img4	2107.276	193.986	14.894	25.253	0.865	0.966	0.176	0.079

Image	MSE Classic	MSE Soft Enhancement	PSNR Classic	PSNR Soft Enhancement	SSIM Classic	SSIM Soft Enhancement	GMSD Classic	GMSD Soft Enhancement
Img5	1787.080	157.079	15.609	26.170	0.890	0.973	0.170	0.070
Img6	2665.781	219.258	13.873	24.721	0.844	0.960	0.200	0.091
Img7	1786.324	121.815	15.611	27.274	0.892	0.975	0.188	0.077
Img8	1906.376	153.911	15.329	26.258	0.883	0.969	0.184	0.069
Img9	2116.673	156.662	14.874	26.181	0.866	0.968	0.193	0.082
Img10	1682.880	136.194	15.870	26.789	0.896	0.975	0.167	0.082

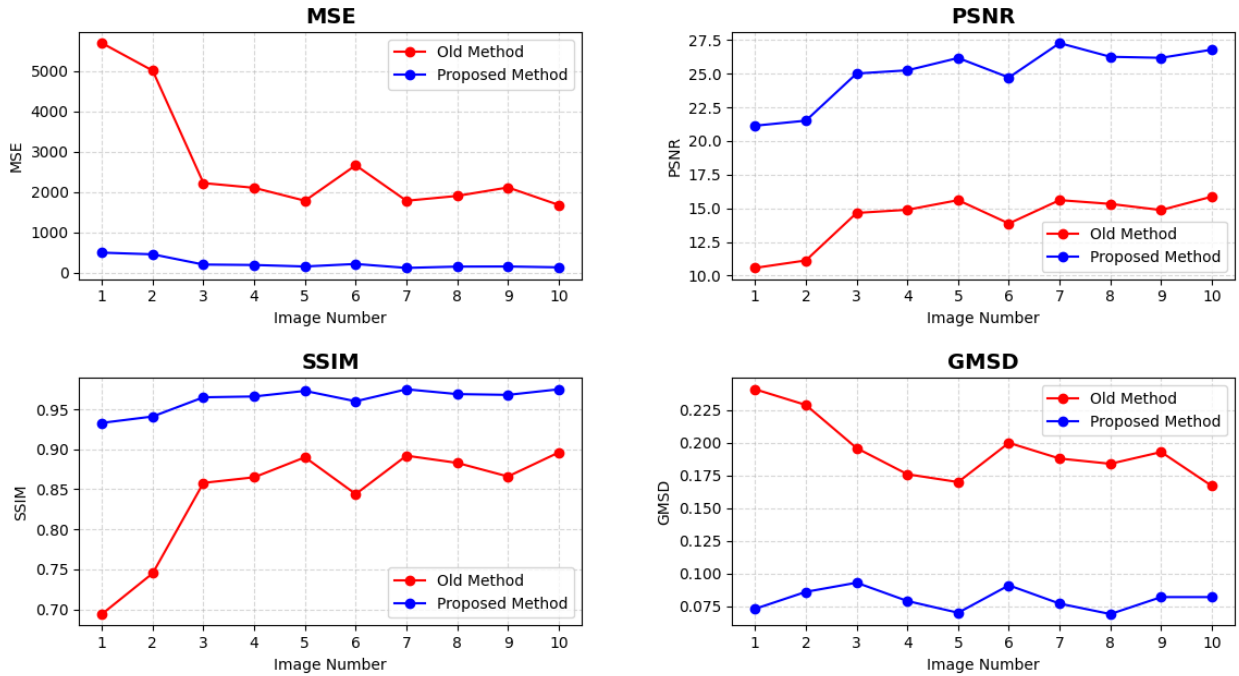


Figure 5. Histogram Comparison of Haze-Reduction Metrics for the Classical and Proposed Methods

8. Discussion

Results from the experiments indicate that the modular two-stage framework suggested has a major positive effect on the quality of scanned images with degradation. The two-stage methodology, by separating shadow removal from haze and blur removal, allows for much greater robustness against the multiple types of degradation compared to a single-stage method. All of the preprocessing techniques (i.e., median filtering and illumination normalization) have provided DRUNet with stable input data, allowing for effective removal of remaining shadows without damaging text features. The improvement in the SSIM scores, as well as the reduction of the GMSD scores, reflect an increase in preservation of document structure and distinct text edges.

Evidence of the efficacy of the shadow removal process can also be seen in the results from the histogram analysis. The use of the shadow removal method results in a more balanced distribution of pixel intensities and better separation between foreground text and background. Improved readability of documents has significant downstream benefits for applications such as optical character recognition (OCR) and document binarization. The results also support the notion that combining classical preprocessing techniques and artificial intelligence algorithms has allowed a system to handle complex illumination changes while maintaining fine detail text information.

The haze and blur reduction modules are also important factors in producing a quality final image. This method enhances local contrast and sharp text without the use of any artefacts by combining Gaussian smoothing, CLAHE, and controlled sharpening together in a single algorithm. The quantitative values indicate a consistent improvement in the metrics of MSE, PSNR, SSIM, and GMSD across the various images tested, indicating that the modularized pipeline provides superior visual/structural fidelity as compared to traditional methods. Overall, these results demonstrate the ability of deep neural networks combined with lightweight enhancement processes to restore scanned documents effectively.

9. Conclusion

An enhancement framework for documents was presented that is implemented in a modular manner to handle two of the most common degradations found in scanned documents: shadowing and haze/blur. The solution does not use one enhancement process but has two distinct enhancement steps, which are complementary to each other. The first step utilizes classical pre-processing methods with a DRUNet-based refinement path to consolidate unevenly illuminated areas and resolve complex shadow structures without compromising the fine structures present within the text. The second step utilizes a soft enhancement technique that uses contrast enhancement and controlled sharpening to reduce haze and blur while preserving the text edge and readability.

In the experimental evaluation of the proposed methodology, the results demonstrate that the hybrid enhancement method is more effective at enhancing clarity and consistency of a document than traditional document enhancement approaches. While traditional filtering processes may struggle with complex types of degradation, they often apply excessive smoothing or add unwanted noise to the document. The combination of a deep learning-based enhancement framework and lightweight enhancement processing steps will allow the proposed enhancement algorithms to restore illumination balance, enhance the text edge, and preserve the quality of key structural features. By employing the above two enhancement methodologies, enhancements will create improved foregrounds and backgrounds and produce a visually enhanced document for use in subsequent processing.

The key insight of this work is that treating shadow removal and haze reduction as two separate yet coordinated processes significantly improves the quality of document restoration. The given framework offers a feasible and computationally effective way of improving scanned documents of poor quality and making them more suitable for digitization and optical character recognition software. This can be extended in future work with more advanced models of restoration, adaptive parameter selection, and testing on larger and more diverse datasets of documents to make it more robust and generalized.

10. Acknowledgements

The authors would like to thank the University of Mosul / College of Computer Science and Mathematics for their facilities, which have helped to enhance the quality of this work.

11. Conflict of Interest

The authors declare that there is no conflict of interest.

12. Funding

The authors declare that they have no known competing financial interests in this paper.

13. REFERENCES

- [1] B. Wang, Z. Wang, W. Liu, X. Huang, C. L. P. Chen, and Y. Zhao, "DDSR-net: Direct document shadow removal leveraging multi-scale attention," *Mach. Intell. Res.*, vol. 22, no. 3, pp. 452–465, 2025, doi: <https://doi.org/10.1007/s11633-024-1522-4>.
- [2] W. Liu, B. Wang, J. Zheng, and W. Wang, "Shadow removal of text document images using background estimation and adaptive text enhancement," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5. doi: DOI: 10.1109/ICASSP49357.2023.10096115.
- [3] M. Guijarro, J. Bayon, D. Martel-Carabias, and J. Recas, "A Multi-Stage Method for Logo Detection in Scanned Official Documents Based on Image Processing," *Algorithms*, vol. 17, no. 4, p. 170, 2024, doi: <https://doi.org/10.3390/a17040170>.
- [4] P. Dubey and D. R. Shashikumar, "A Multi-Stage Deep Learning Model for the Enhancement of the Quality of Camera-Captured Document Images," *Eng. Technol. & Appl. Sci. Res.*, vol. 15, no. 6, pp. 29964–29970, 2025, doi: <https://doi.org/10.48084/etasr.14469>.
- [5] D. Gao, W. Liu, S. Chen, J. Qiu, X. Mei, and B. Wang, "Document Image Shadow Removal Based on Illumination Correction Method," *Algorithms*, vol. 18, no. 8, p. 468, 2025, doi: <https://doi.org/10.3390/a18080468>.
- [6] Y. Jin, Y. Wang, Q. Zhong, K. C. Jin-Chun, K. Z. Ke, and D. MacDonald, "Multi-Stage Field Extraction of Financial Documents with OCR and Compact Vision-Language Models," *arXiv Prepr. arXiv2510.23066*, 2025, doi: <https://doi.org/10.48550/arXiv.2510.23066>.

- [7] Z. Anvari and V. Athitsos, "A survey on deep learning based document image enhancement," *arXiv Prepr. arXiv2112.02719*, 2021, doi: <https://doi.org/10.48550/arXiv.2112.02719>.
- [8] M. Karayaka, U. Muhammad, J. Laaksonen, M. Z. Hoque, and T. Seppänen, "A Dual-Domain Convolutional Network for Hyperspectral Single-Image Super-Resolution," *arXiv Prepr. arXiv2512.09546*, 2025, doi: <https://doi.org/10.48550/arXiv.2512.09546>.
- [9] B. Wang, C. Li, W. Zou, Y. Zhang, X. Chen, and C. L. P. Chen, "A comprehensive survey on shadow removal from document images: datasets, methods, and opportunities," *Vicinagearth*, vol. 2, no. 1, p. 1, 2025, doi: <https://doi.org/10.1007/s44336-024-00010-9>.
- [10] Y. Cui, Y. Tao, L. Jing, and A. Knoll, "Strip attention for image restoration," in *International Joint Conference on Artificial Intelligence, IJCAI*, 2023. doi: <https://doi.org/10.24963/ijcai.2023/72>.
- [11] C. Wang *et al.*, "Promptrestorer: A prompting image restoration method with degradation perception," *Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 8898–8912, 2023, doi: <https://dl.acm.org/doi/abs/10.5555/3666122.3666512>.
- [12] Z. Chen, Z. He, and Z.-M. Lu, "DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *IEEE Trans. image Process.*, vol. 33, pp. 1002–1015, 2024, doi: DOI: 10.1109/TIP.2024.3354108.
- [13] X. Chen *et al.*, "Unpaired deep image dehazing using contrastive disentanglement learning," in *European conference on computer vision*, 2022, pp. 632–648. doi: DOI https://doi.org/10.1007/978-3-031-19790-1_38.
- [14] G. L. N. Vanguri, S. K. Swain, and M. V. Krishna, "An Investigative Study on Deep Learning-Based Image Dehazing Techniques," in *International Conference on Computational Innovations and Emerging Trends (ICCIET-2024)*, 2024, pp. 87–97. doi: DOI 10.2991/978-94-6463-471-6_9.
- [15] Y. Zhou, S. Zuo, Z. Yang, J. He, J. Shi, and R. Zhang, "A review of document image enhancement based on document degradation problem," *Appl. Sci.*, vol. 13, no. 13, p. 7855, 2023, doi: <https://doi.org/10.3390/app13137855>.
- [16] Z. Zhou *et al.*, "Docdeshadower: Frequency-aware transformer for document shadow removal," in *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2024, pp. 2468–2473. doi: DOI: 10.1109/SMC54092.2024.10831480.
- [17] B. Wang and C. L. P. Chen, "Local water-filling algorithm for shadow detection and removal of document images," *Sensors*, vol. 20, no. 23, p. 6929, 2020, doi: <https://doi.org/10.3390/s20236929>.
- [18] X. Chen, X. Cun, C.-M. Pun, and S. Wang, "Shadocnet: Learning spatial-aware tokens in transformer for document shadow removal," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5. doi: DOI: 10.1109/ICASSP49357.2023.10095403.
- [19] A. Fathallah, M. El Yacoubi, and N. Ben Amara, "EHDI: enhancement of historical document images via generative adversarial network," in *18th International Conference on Computer Vision Theory and Applications (VISAPP)*, 2023, pp. 238–245. doi: DOI : 10.5220/0011662700003417.
- [20] H. Neji, T. M. Hamdani, M. B. Halima, J. Nogueras-Iso, and A. M. Alimi, "Blur2sharp: A gan-based model for document image deblurring," 2021. doi: DOI: 10.2991/IJCIS.D.210407.001.
- [21] Z. Yang *et al.*, "Docdiff: Document enhancement via residual diffusion models," in *Proceedings of the 31st ACM international conference on multimedia*, 2023, pp. 2795–2806. doi: <https://doi.org/10.1145/3581783.3611730>.
- [22] A. A. Abdulmajeed and N. N. Saleem, "Pre-processing and enhancement techniques for COVID-19 X-ray and CT-scan medical images," in *AIP Conference Proceedings*, 2022, p. 50011. doi: <https://doi.org/10.1063/5.0121956>.
- [23] N. T. Saeed and H. M. Ahmed, "Building a Real-Time System to Monitor Students Electronically Based on Digital Images of Face Movement," *ICOASE 2022 - 4th Int. Conf. Adv. Sci. Eng.*, pp. 83–88, 2022, doi: 10.1109/ICOASE56293.2022.10075587.
- [24] H. S. Abdulla, A. S. Shaheen, and N. M. Isaac, "Effectiveness of Image Curvelet Transform Coefficients for Image Denoising", doi: <https://doi.org/10.33899/csmj.2024.146534.1105>.
- [25] J. Xu *et al.*, "General information metrics for improving AI model training efficiency," *Artif. Intell. Rev.*, vol. 58, no. 9, p. 289, 2025, doi: <https://doi.org/10.1007/s10462-025-11281-z>.
- [26] M. D. Badiuzzaman, "Unpacking the metrics: a critical analysis of the 2025 QS World University Rankings using Australian university data," in *Frontiers in Education*, 2025, p. 1619897. doi: <https://doi.org/10.3389/educ.2025.1619897>.
- [27] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, 2021, doi: DOI: 10.1109/TPAMI.2021.3088914.
- [28] R. Vimal, A. Kumar, and A. Tiwari, "A novel compression method for Vectorcardiogram signal using hybrid autoencoder model," *Biomed. Signal Process. Control*, vol. 113, p. 108796, 2026, doi: <https://doi.org/10.1016/j.bspc.2025.108796>.
- [29] A. Dziembowski, W. Nowak, and J. Stankowski, "IV-SSIM-The structural similarity metric for immersive video," *Appl. Sci.*, vol. 14, no. 16, p. 7090, 2024, doi: <https://doi.org/10.3390/app14167090>.
- [30] A. Di Marino, V. Bevilacqua, E. Di Nardo, A. Ciaramella, I. De Falco, and G. Sannino, "A new Image Similarity Metric

for a Perceptual and Transparent Geometric and Chromatic Assessment,” *arXiv Prepr. arXiv2601.19680*, 2026, doi: <https://doi.org/10.48550/arXiv.2601.19680>.

تحسين المستندات المسووحة ضوئياً باستخدام المعالجة المعيارية متعددة المراحل

زهراء علي محمد نضير^(1*)، حسن ماهر احمد⁽²⁾

^(1,2) قسم البرمجيات، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق

المستخلص

غالبًا ما تكون صور المستندات المسووحة ضوئياً عرضةً لتشوهات معقدة، بما في ذلك الإضاءة غير المتساوية والظلال والضباب، مما يقلل بشكل كبير من وضوح النص وأداء أنظمة التعرف الضوئي على الحروف (OCR). تقدم هذه الورقة نموذجًا معياريًا متعدد الخطوات يهدف إلى تحسين جودة المستندات من خلال معالجة كل تشوه من هذه التشوهات على حدة. يتألف تصميم النظام من وحدتين محددين: وحدة إزالة الظلال ووحدة تقليل الضباب/التشويش. تستخدم خوارزمية إزالة الظلال تقدير الخلفية باستخدام الترشيح الوسيط وتطبيع الإضاءة، ثم خطوة تحسين باستخدام نموذج DRUNet المدرب مسبقًا لتجنب الضوضاء المتبقية، على حساب دقة تفاصيل النص. بالإضافة إلى ذلك، يستخدم نظام تقليل الضباب تقنية تحسين ناعمة تعتمد على معادلة التباين التكميلية المحدودة (CLAHE) مع التنعيم الغاوسي لتحقيق تحسين موضعي للتباين دون أي تشوهات بصرية. تشير النتائج التجريبية إلى أن النهج المعياري المقترح يُنتج خلفية أكثر وضوحًا وحوافًا أكثر دقة في النص مقارنةً بالطريقة التقليدية أحادية المرحلة. تُعدّ قياسات MSE و PSNR و SSIM و GMSD تحليلات كمية لفعالية هذا الإطار، وتُظهر النتائج تحسنًا مائلًا في جميع أنواع المستندات.