Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.18.(1) 2026,pp.Comp 160–172

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Designing a Modular Framework for Processing and Enhancing Scanned Documents Using Advanced Denoising Algorithms

Zahraa Ali Mohamed Nather [a, *], Hasan Maher Ahmed [a]

[a] *Software Department, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq. Email: zahraa.24csp18@student.uomosul.edu.iq, hasanmaher@uomosul.edu.iq*

ARTICLE INFO

Article history:
Received: 09 /11/2025
Rrevised form: 05 /01/2026
Accepted : 12 /01/2026
Available online: 30 /03/2026

Keywords: Scanned Documents, Image Processing, Document Enhancement, Denoising Algorithms, Modular Framework

ABSTRACT

Scanned documents are often flawed such as background noise, skew and uneven illuminations which negatively affect reading and text recognition. The given study presents a two-step processing model that can serve to improve the quality of grayscale and color-scanned documents because of the use of combined denoising and deskewing methods. Three denoising methods were tested under various noise levels: DRUNet, DnCNN, and Total Variation (TV). To get the best results in restoring image quality, we used pre-trained models for the deep learning algorithms (DRUNet and DnCNN) through the DeepInv library. This allowed us to use powerful, ready-to-use features to clean the documents effectively. Their performance was then measured using standard quality scores and visual checks. The results showed that DRUNet produced the best and more reproducible performance, which was able to suppress noise and preserve fine structural and textual fidelity. In addition, preprocessing step of Otsu thresholding and minimum bounding rectangle estimation was also applied to automatically correct document skew to enhance text alignment and readability. Python and Gradio were used as the implementation language of the system to offer an interactive, transparent, and reproducible platform. In general, the suggested framework would significantly increase the clarity, alignment, and the overall quality of scanned documents and make them more reliable to use in OCR and digital archiving purposes.

MSC: 68U10; 68T05; 68T07; 68W10

https://doi.org/10.29304/jqcsm.2026.18.12441

1- Introduction

The scanned documents are one of the most crucial types of digital information in the modern world. They serve as online equivalents of paper-based records, such as bibliographic content, government records, archives, and government files. Through digitization through scanning, priceless information is also preserved, shared, and processed electronically hence making them easily stored as well as granting them long-term accessibility.

∗Corresponding author: Zahraa Ali Mohamed Nather

Email addresses: zahraa.24csp18@student.uomosul.edu.iq

Communicated by ‘sub etitor’

The importance of scanned documents lies in the central role of the scanned document in the digitization, automation, and information retrieval systems. They provide a practical system through which organizations, libraries, and institutions can convert large amounts of paper-based content to searchable and easily manageable digital forms, ensuring that important data is not destroyed through physical means and that it can then be used in modern applications, such as optical character recognition (OCR) and document indexing, and digital archiving.

Yet they are very important, the scanned documents are often found to possess various defects which are the result of the processes used in scanning. Some of the defects most frequently met with are image skew, noise, bad illumination and deterioration or damage of paper, which cause loss of quality. Such defects may greatly impair the value of the documents in that they decrease the legibility very materially, and decrease the accuracy of various subsequent operations on the documents such as reading or extracting the text, classifying and retrieving the material. Recent studies have highlighted that correcting skew (tilt) in scanned documents is a crucial step to improve the usability and accuracy of subsequent processing [1]. So, the improvement of the documents as far as possible so that the legibility and registration of the documents may be improved is one of the most important steps to be taken in the direction of setting up an efficient digital preservation and capable recognition process. Studies have also shown that using deep learning methods for denoising scanned documents can effectively restore image quality, even without access to clean reference images [2].

This paper presents a two-stage improvement operation which may be used to overcome some of these defects and includes deskewing and denoising methods on both gray scale and color documents. The system essentially uses the deep learning methods to clean up the images and recover their visual fidelity. In addition, it automatically compensates skew or rotational distortion before other processing operations take place. In addition, a Python/Gradio based interactive platform was created enabling real time experimentation and reproducibility. All in all, the given work will help to enhance the clarity, format, and functionality of scanned documents, making a good base of proper OCR and digital archiving systems. -

2-LITERATURE REVIEW

Recent advances in document image processing have provided enhanced methodologies leading to optimized performance of OCR. Old principles determined through geometry and statistical processes do not dominate research on processes; the application of deep learning principles is frequently used in these areas because they specifically address important principles of noise, skew and misplacement of text. This is a review of important research developed on enhancement, denoising and post-processing with respect to OCR and deals specifically with the problems existing in the area.

Rotman et al. designed a framework for U-Net processes in preprocessing adding modules of denoising and enhancement which are trained on synthetically noisy document datasets which enhances the resilience and performance of optical character recognition systems. The proposed solution is sufficiently efficient to handle the various aspects of the totality of the type of noising and degrading effects to which a document is subjected resulting in more robust performance and generally consistent OCR results across a wide range of datasets [3].

Boudraa et al. Proposed a method for determination of skew for historical documents which applies skeletonisation processed with modified Hough transforms applied to determine the skew of the text orientation. The method used is effective in dealing with correcting geometric illusions and enhances performance for OCR and legibility improvements with degradation or age of the document images and is effective for OCR improvement [4].

Shafi et al. Created and designed a hybrid framework which is an application of tilt correction as a function of the Hough Transform model coupled with a lightweight deep learning model enabling adequate correction of tilt and inversion errors thus utility correcting skew and inversion errors in documents thus. This is a preprocessing methodology which allows improved form for text placement and thus orientation of text and amplifies the accuracy and performance of the other OCR processing [5].

Kim et al. performed a systematic review of deep learning-based image denoising techniques, providing a synthesis of the different architectural paradigms, algorithmic techniques, and areas of use. The article highlights the effectiveness of neural network-based techniques in reducing noise in images whilst maintaining the salient image features, and outlines their future potential in diverse applications, such as in medical imaging [6]. Zangana et al. suggested a hybrid approach that involves a convolutional neural network and a wavelet transform-based denoising algorithm. This method is useful in reducing noise as well as maintaining critical structural and textual information of document images. It makes both printed and handwritten text more legible and readable by maintaining edges and fineness. This results in the approach enhancing the performance and reliability of the later OCR processing of noisy or degraded documents [7].

Tawfik et al. compared traditional and deep learning-based denoising models in terms of their applications on micro-computed tomography images. The paper explains the advantages and disadvantages of each method, providing insights into the most effective techniques for enhancing image quality without compromising important structural features [8].

Abdulla et al. suggested the use of 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 [9].

Some of the most groundbreaking hybrid filtering techniques that were proposed by Ali are for color image denoising by a combination of a mean and conservative smoothing algorithm. This method is efficient in smoothing noise without any loss of significant color information and edges to improve the quality and clarity of the images to be further processed and analyzed [10].

Supriyono et al. performed a systematic review that combines natural language processing (NLP) and deep learning methods in order to perform post-processing of the results of optical character recognition (OCR) and in order to perform automatic text summarization. The research shows that a hybrid of the two methods improves the quality of the OCR output and makes it possible to generate summaries of information in a concise manner, thus increasing the efficiency of the process of extracting information in scanned or other digitized documents [11].

Hukkeri et al. conducted a benchmark study to compare OCR techniques on various types of media, including frames extracted from lecture videos. The study assessed the robustness, accuracy, and performance of different OCR methods, providing insights into the best approaches for handling complex and varied visual content [12]

Taken together, these works highlight the ongoing development in improved OCR and post-correction, while deep learning still serves as a driving force to solve existing problems of noise artifacts, ambiguous context of detection regions, and variability of text along languages or writing scripts.

3- DATASET

The experimental evaluation was carried out on a self-curated dataset of forty high-resolution scanned images collected in 2025.These scans were only obtained in academic lectures and university archives to get challenges that are associated with documents; the collection was equally divided into twenty color images and twenty grayscale images. All the pictures were taken with 300 dots/in. spatial resolution to provide high-fidelity images.

The strength of the framework was tested based on the original scans as the ground truth which was then degraded by introducing the synthetic Gaussian noise in three preset standard deviations ($\sigma = 0.05, 0.10, 0.20$). In turn, this dataset provides a heterogeneous benchmark, which enables the comparison of the denoising and deskewing algorithms with a wide range of types of documents.

4- METHODOLOGY

This method has simple steps that are clear where 3 algorithms are put into tests and comparisons with others: TV, DnCNN and DRU-net. Process starts with preparation of the data, and added intentionally artificial noise is added. From this point these algorithms are put into practice to the images. Finally, the evaluations of these since each algorithm is responsible for enhanced quality of images.

4.1 Model Architecture

This method follows a few basic steps shown in Figure 1. First, data is prepared using scanned document images. Then, noise is added at different levels. Three different algorithms (TV, DnCNN, and DRUNet) are used to remove the noise. After that, the algorithms are tested using numbers and also by how people see the images. Finally, the results are shown clearly and compared to help understand them better.

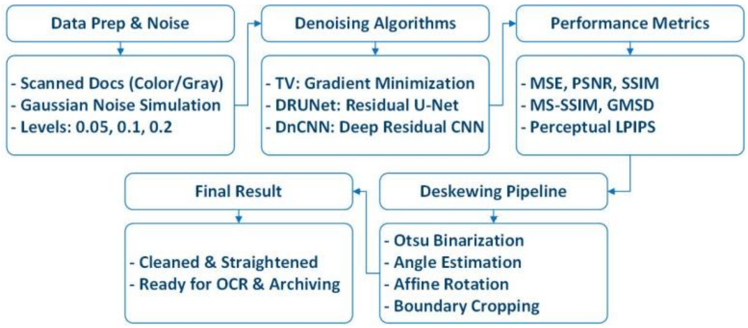

Fig. 1. Proposed methodology architecture for image denoising.

4.2 Denoising Stage: Image Noise Reduction

Three algorithms (TV, DRUNET and DnCNN) were used to remove noise from color and grayscale images in order to improve the quality of scanned images, making them clearer and sharper.

4.2.1. Total Variation (TV)

Is a model-driven denoising algorithm that controls the gradient of the image by minimizing the distance between pixels. The method is especially effective in the even-ground areas and allows maintaining the sharpness of the textual and structural details. The TV denoiser was applied to both grey scale and color imagery to ensure that the noise is reduced without dislocation of the necessary textures [13].

Algorithm: Classical Total Variation (TV) Denoising

Step 1: Problem Formulation given a noisy image y, the goal is to recover the clean image x by solving the following optimization problem:

$$\min_{x} \frac{1}{2} \| x - y \|_2^2 + \lambda\, TV(x) \tag{1}$$

Step 2: Total Variation Definition depending on the image characteristics, TV can be defined in two ways:

Isotropic TV:

$$TV(x) = \sum_{i,j} \sqrt{(D_x x_{i,j})^2 + (D_y x_{i,j})^2} \tag{2}$$

Anisotropic TV:

$$TV(x) = \sum_{i,j} \left(|D_x x_{i,j}| + |D_y x_{i,j}| \right) \tag{3}$$

Step 3: Euler-Lagrange Equation the optimality condition for the TV minimization is derived from the Euler-Lagrange equation:

$$x - y - \lambda \nabla \cdot \left(\frac{\nabla x}{|\nabla x|} \right) = 0 \tag{4}$$

Step 4: Iterative Solution the solution is typically found using an iterative approach such as gradient descent:

$$x^{(k+1)} = x^{(k)} + \tau \left[y - x^{(k)} + \lambda \nabla \cdot \left(\frac{\nabla x^{(k)}}{|\nabla x^{(k)}| + \epsilon} \right) \right] \tag{5}$$

Step 5: Parameter Selection the regularization parameter $\lambda$is critical as it controls the degree of smoothing:

- **Large λ:** results in strong smoothing but may lead to a loss of fine details.
- **Small λ:** provides weak smoothing, potentially leaving residual noise in the output.

Step 6: Output the final denoised image $\hat{x}$ is obtained after convergence of the iterative process.

4.2.2. DRUNet (Deep Residual U-Net):

DRUNet is a state-of-the-art deep learning model to image denoising that was proposed by Zhang et al. It incorporates a strong U-Net architecture with residual learning, in such a way that the network is trained to learn the difference between the noisy and clean images, but not the clean image DRUNet uses skip connections, and a deep denoiser before preserving fine details and is able to effectively reduce strong structured noise. This architecture is found to have high performance at varying level of noises and it keeps image fidelity and hence can be used in medical imaging, satellite imagery and also industrial inspection. The flexibility and the capability of reassembling finer structures makes DRUNet an effective part of both color and grayscale image tasks [14].

In this work, a pre-trained DRUNet model was employed without retraining, following the standard configuration proposed by the original authors.

Algorithm**: DRUNet (Deep Residual U-Net)**

Step 1: Noise Model (AWGN) We assume the input document y is degraded by Additive White Gaussian Noise (AWGN)

$$y = x + n, \quad n \sim N(0, \sigma^2) \tag{6}$$

Step 2: Residual Learning Formulation The core innovation is that the network learns to estimate the noise residual $\mathcal{R}(y)$instead of the clean image x:

$$\mathcal{R}(y) = \mathcal{F}(y; \Theta) \approx y - x \tag{7}$$

Where $\mathcal{F}$ represents the residual mapping function, and Θ denotes the trainable parameters of the model.

Step 3: U-Net Hierarchical Architecture The architecture utilizes an encoder-decoder structure with skip connections to capture both high-level semantic features and low-level spatial details:

Encoder: $z_l = \mathrm{Enc}(z_{l-1})$, Bottleneck: $z_{mid} = \mathrm{Conv}(z_L)$, **Decoder**: $\hat{z_l} = \mathrm{Dec}(\widehat{z_{l+1}}, z_l)$

Step 4: Image Reconstruction (Output) The clean document x is reconstructed by subtracting the estimated residual from the original noisy input:

$$\hat{x} = y - \mathcal{R}(y) \tag{8}$$

Step 5: Optimization Objective (Loss Function) During the training phase, the model parameters Θ were optimized by minimizing the Residual Mean Squared Error (MSE):

$$\mathcal{L}(\Theta) = \frac{1}{N}\sum_{i=1}^{N} |\mathrm{Net}(y_i, \sigma_i; \Theta) - (y_i - x_i)|^2 \tag{9}$$

4.2.3. DnCNN: Deep Convolutional Neural Network

which is image denoising. It uses residual learning to understand the noise of an image, and not the clean image itself, which enhances the performance of denoising and efficiency. The batch normalization is to stabilize and speed up the training. DnCNN is also very effective in coping with Gaussian noise of known or unknown levels and can be generalized to other problems such as super-resolution and JPEG deblocking, where traditional approaches always do poorly in quality and speed [15].

Algorithm: DnCNN (Denoising Convolutional Neural Network)

Step 1: Input and Noise Modeling The observed noisy image y is modeled as the sum of the latent clean image x and the additive noise n:

$$y = x + n \tag{10}$$

Where x: clean image, n: noise, y: observed noisy image

Step 2: Feature Extraction (First Layer) The network begins by extracting initial features through a series of convolutional filters combined with a Rectified Linear Unit (ReLU) activation

$$z_1 = \mathrm{ReLU}\big(\mathrm{Conv}(y)\big)F_1 = \sigma(W_1 * y + b_1) \tag{11}$$

Step 3: Feature Refinement and Normalization (Intermediate Layers) For layers l = 2 to D-1, the model applies Convolution, Batch Normalization (BN), and ReLU to refine the features:

$$z_l = \mathrm{ReLU}\left(\mathrm{BN}\big(\mathrm{Conv}(z_{l-1})\big)\right) \tag{12}$$

Step 4: Residual Noise Estimation (Final Layer) The final layer outputs the estimated noise residual $\mathcal{R}(y)$, isolating the degradation from the structural content:

$$\mathcal{R}(y) \approx \eta \tag{13}$$

Step 5: Image Reconstruction The clean document is recovered by subtracting the estimated residual noise from the original noisy input:

$$\hat{x} = y - \mathcal{R}(y) \tag{14}$$

Step 6: Loss Function (Optimization Objective) During training, the parameters Θ are optimized using the Mean Squared Error (MSE) between the predicted and true residual:

$$\mathcal{L}(\Theta) = \frac{1}{2N}\sum_{i=1}^{N} |\mathcal{R}(y_i;\Theta) - (y_i - x_i)|^2 \tag{15}$$

Step 7: Parameter Update The model weights are updated via stochastic gradient descent or its variants to minimize the objective function:

$$\Theta \leftarrow \Theta - \alpha\nabla\mathcal{L}(\Theta) \tag{16}$$

3.4 Skew Detection and Correction (Color and Grayscale Images)

In this step, skew detection and correction of all the scanned document images (color and grayscale) were performed to make sure that there was correct positioning after using the denoising algorithms. Originally there may be an artificial skewness to model imaginary real world scanning errors. The individual pictures were then converted to the correct format that can be analyzed and the skew angle was determined by taking advantage of pixel intensity and contour data. After the skew angle had been determined, the pictures were rotated to put them in their right orientation. Lastly, white space extraneous of content was cut out to keep only the content. This preprocessing pipeline therefore ensures that the images are appropriately positioned, purified and prepared to undergo other denoising algorithms [1].

4.3.1. Algorithm: Deskewing (Color and Grayscale Images)

Step 1: Image Binarization (Otsu's Method) To achieve the highest accuracy in angle detection, the image is converted to a binary format to isolate text components from the background. We apply Otsu's Method to automatically determine the optimal threshold τ:

$$I_{Binary}(i,j) = \begin{matrix} 1 \text{if } I_{Gray}(i,j) < \tau \\ 0 \text{otherwise} \end{matrix} \tag{17}$$

Step 2: Skew Angle Estimation We identify the coordinates P of all text pixels and fit a minimum area bounding rectangle around them to extract the precise skew angle θ:

$$\theta = MinAreaRect(P) \rightarrow \text{angle} \tag{18}$$

Step 3: Rotation (Affine Transformation) The original image is rotated by the inverse of the detected angle θ to restore horizontal alignment. This is mathematically executed using a rotation matrix:

$$[i'j'] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} [ij] \tag{19}$$

Step 4: Boundary Refinement (Cropping) Post-rotation, redundant white margins are generated at the corners. The image is cropped based on the bounding box of the actual content to ensure a clean final output:

$$I_{Final} = \text{Crop}(I_{Rotated}, \text{BoundingBox}) \quad (20)$$

5-IMPLEMENTATION

The system was developed in Python 3.10 on Google Colab and utilized GPU for faster image processing provided to the users more responsive, especially large resolution scanned images. The key libraries that have been used are: OpenCV to read and convert images into grayscale, fix tilts of the photos and apply a threshold; NumPy for reverse calculations; PIL for opening and saving images; DeepInv helped to implement denoising models such as DnCNN, DRUNet, or TV denoiser. The first step is the denoising of the image. All images are resized to 512×512, and converted to gray-scale if necessary. We then deliberately introduce some Gaussian noise in the images to make them more realistic noisy images. Finally, one of the denoising models is applied to remove. All the intermediate and final results are automatically saved for quantitative metric evaluation as well as visual inspection. This cascaded mode allows us to perform the denoising first and then deskewing, which makes the skew angle computation more precise and prevents noise influence when angle is computed.

6.RESULT

6.1 Results on Grayscale and Color Scanned Images:

A qualitative comparative study was performed on the practical effectiveness of the denoising algorithms to assess the impact of the algorithm on the Gaussian noise as the main degradation model. Gaussian noise with varying noise levels of 0.05, 0.1, and 0.2 were used to recreate the electronic and thermal noise present in the scanning sensors with high-resolution giving the standardized benchmark on noise reduction of the scanning sensor. Fig. 2 below depicts the performance of TV, DnCNN and DRUNet in terms of restoring fine textual content and preserving fidelity of both grayscale and color document images. Qualitative performance of TV, DnCNN, and DRUNet algorithms on a grayscale (Top) and color (Bottom) document images at varying levels of Gaussian noise levels (0.05, 0.1, and 0.2). In both types of images, the subfigure (a) represents the reference; (b-d) represents noisy inputs; (e-g) represents TV; (h-j) represents DnCNN; and (k-m) represents DRUNet. As can be seen by the visual results in Fig. 4, DRUNet remains outperforming the other methods because of its hybrid U-Net/ResNet architecture and adding a noise level map. This design is adaptive in balancing noise reduction and edge preservation enabling such a design to retain high-frequency textual information and structural integrity. On the contrary, TV triggers staircasing artifacts and DnCNN is prone to over-smoothing, DRUNet is the most resistant solution to grayscale and color document restoration.

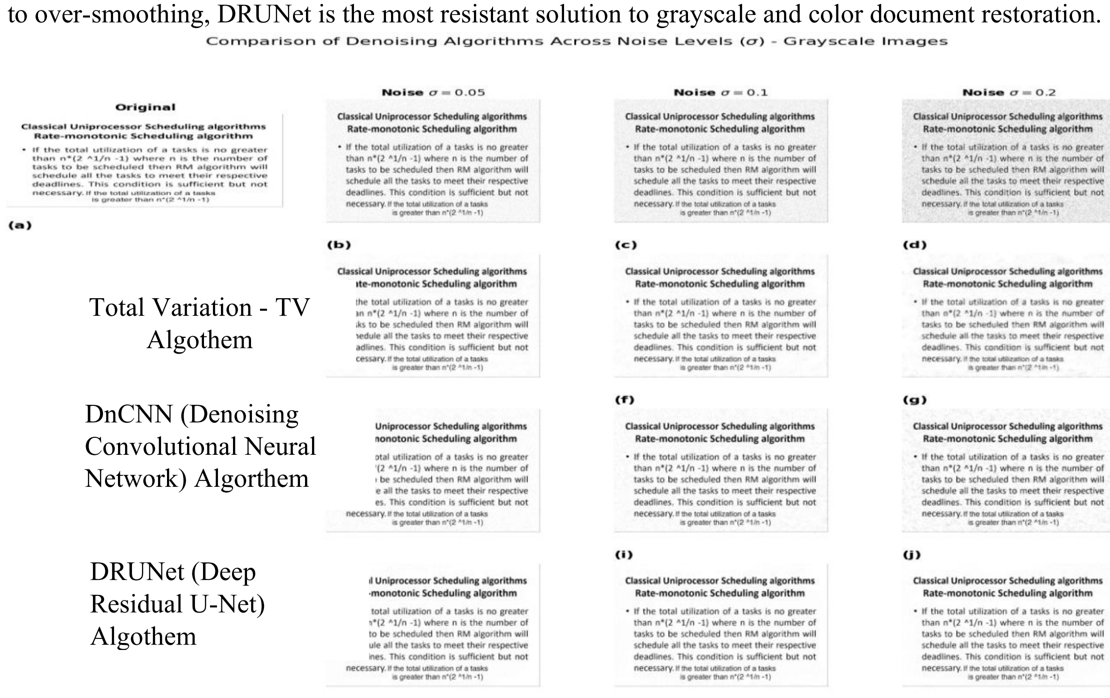

Fig. 2: Visualization of denoising algorithm on grayscale scanned documents at different levels of noise (0.05, 0.1, 0.2).

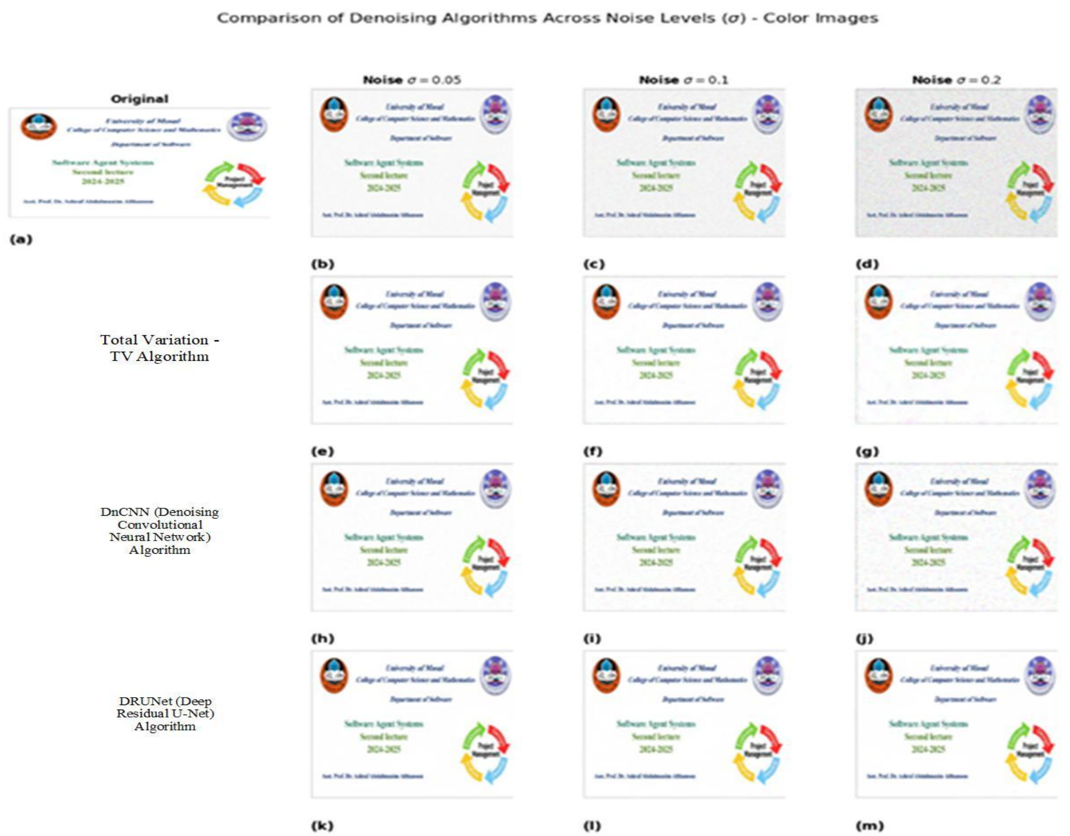

Fig. 3: Visualization of denoising algorithm on color scanned documents at different levels of noise levels) 0.05, 0.1, 0.2).

6-3 Skew Correction Results in Scanned images.

This section demonstrates that the frameworks' ability to detect and rectify misalignments in both grayscales scanned documents and color documents. Achieving precise orientation is received and structural regularity is brought into which the visual quality and additional processing of the visual image is enhances.

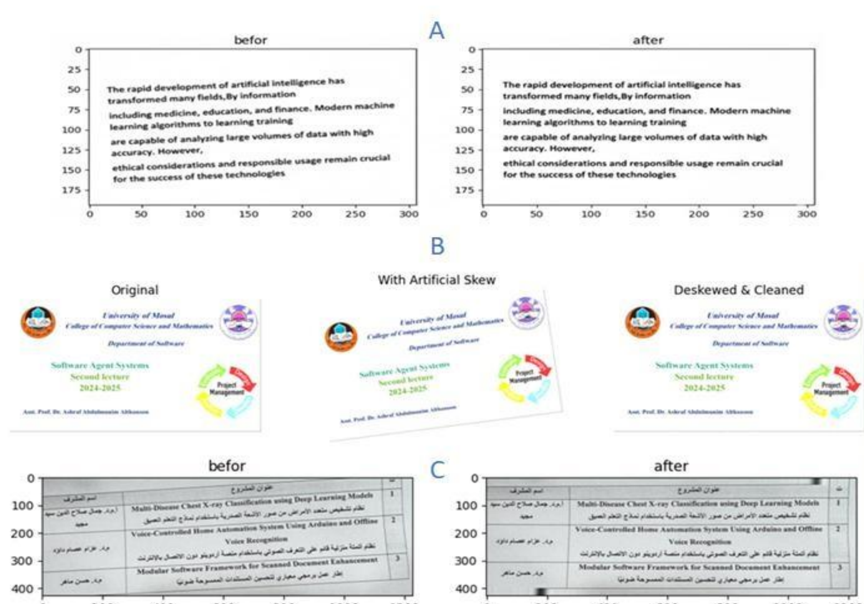

Fig.4: A: Deskewing results for grayscale documents, B: color scanned images, C: Deskewing results for grayscale

7. Evaluation Metrics

Six popular metrics of image quality assessment (IQA) that included Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Multi-Scale Structural Similarity Index (MS-SSIM), Learned Perceptual Image Patch Similarity (LPIPS) and Gradient Magnitude Similarity Deviation (GMSD) were used to evaluate the denoising performance.

7.1.1. Mean Squared Error (MSE)

MSE measures the average squared difference between the reference image I and the denoised image $\tilde{I}$, quantifying pixel-wise reconstruction accuracy:

$$\text{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j) - \hat{I}(i,j)]^2 \tag{21}$$

A lower MSE indicates higher restoration accuracy and less distortion [6]. Recent studies continue to use MSE as a baseline for evaluating denoising performance.

7.1.2. Peak Signal-to-Noise Ratio (PSNR)

PSNR expresses the ratio between the maximum possible pixel intensity I_{max} and the MSE, measured in decibels (dB):

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{I_{max}^2}{\text{MSE}}\right) \tag{22}$$

The higher value of PSNR tends to suggest an improved quality of visuals [16]. It is a simple and strong quantitative evaluation broadly applied to the notion of image restoration performance.

7.1.3. Structural Similarity Index (SSIM)

SSIM evaluates perceptual similarity between images based on luminance (l), contrast (c), and structure (s) components:

$$\text{SSIM}(I,\hat{I}) = [l(I,\hat{I})]^{\alpha} \cdot [c(I,\hat{I})]^{\beta} \cdot [s(I,\hat{I})]^{\gamma} \tag{23}$$

Where:

$$l(I,\hat{I}) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1}, \quad c(I,\hat{I}) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2}, \quad s(I,\hat{I}) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I\sigma_{\hat{I}} + C_3} \tag{24}$$

Values close to **1** indicate high structural similarity and preserved visual perception [17].

7.1.4. Multi-Scale Structural Similarity Index (MS-SSIM)

MS-SSIM extends SSIM by evaluating image similarity over multiple spatial scales:

$$\text{MS-SSIM}(I,\hat{I}) = [l_M]^{\alpha_M}\prod_{j=1}^{M}[c_j]^{\beta_j}[s_j]^{\gamma_j} \tag{25}$$

It captures perceptual details across resolutions, making it highly effective for complex textures and deep-learning-based denoising models [16], [17]

7.1.5. Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS measures perceptual similarity using deep neural network feature embeddings. It computes the distance between feature activations extracted from a pretrained network (e.g., VGG, AlexNet):

$$\mathrm{LPIPS}(I,\hat{I}) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \| w_l \odot (f_l(I)_{hw} - f_l(\hat{I})_{hw}) \|_2^2 \tag{26}$$

Where $f_l(\cdot)$ represents the feature map at layer l, and w_l are learned weights.
Lower LPIPS scores indicate higher perceptual similarity [18]

7.1.6. Gradient Magnitude Similarity Deviation (GMSD)

GMSD assesses image quality by comparing the gradient magnitude similarity (GMS) maps of the reference and test images:

$$\mathrm{GMS}(i,j) = \frac{2G_I(i,j)G_{\hat{I}}(i,j)+C}{G_I(i,j)^2+G_{\hat{I}}(i,j)^2+C} \tag{27}$$

Then, the overall quality score is computed as the standard deviation of the GMS map:

$$\mathrm{GMSD} = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(\mathrm{GMS}(i,j) - \overline{\mathrm{GMS}})^2} \tag{28}$$

Lower GMSD values correspond to higher perceptual consistency and better edge preservation [19], [20]

7.2 Denoising Color and Grayscale Image Denoising.

In this section of the work, the algorithms are rated by six metrics MSE and PSNR used to estimate pixel-level accuracy and SSIM, MS-SSIM, LPIPS and GMSD are used to estimate the perceptual quality and structural integrity as interpreted by the human eye. To interpret the results, we took the mean performance of each algorithm to all the images and one can see that there were 20 color images and 20 grayscale images. The quantitative findings are presented in Tables 1 and 2, and the visual comparisons in Figures 4 and 5 with charts to demonstrate the trends in performance of each algorithm and noise level and finally, what methods proved to be the most efficient and effective.

Table1: Presents the average metrics for the grayscale image set.

Index	Algorithm	Noise Level (σ)	MSE	PSNR (dB)	SSIM	MS-SSIM	LPIPS	GMSD
0	**DRUNet**	0.05	0.0002	**37.8073**	0.9968	0.9988	0.0011	0.0197
1	DnCNN	0.05	0.0004	34.0452	0.9690	0.9917	0.0058	0.0295
2	TV	0.05	0.0019	27.3914	0.9630	0.9893	0.0303	0.0741
3	**DRUNet**	0.1	0.0006	**32.2996**	0.9888	0.9959	0.0038	0.0371
4	DnCNN	0.1	0.0015	28.4499	0.9060	0.9712	0.0275	0.0569
5	TV	0.1	0.0057	22.6558	0.8824	0.9595	0.1146	0.1249
6	**DRUNet**	0.2	0.0022	**26.9104**	0.9613	0.9855	0.0125	0.0677
7	DnCNN	0.2	0.0050	23.1665	0.7606	0.9084	0.1022	0.1060
8	TV	0.2	0.0127	19.1704	0.7184	0.8630	0.3366	0.1780

Table 2: Presents the average metrics for the color image set.

Algorithm	Noise Level (σ)	PSNR (dB)	SSIM	MS-SSIM	LPIPS	GMSD	MSE
DRUNet	0.05	**39.99**	**0.9950**	**0.9987**	**0.0081**	**0.0145**	**0.00011**
	0.1	**36.06**	**0.9911**	**0.9975**	**0.0129**	**0.0223**	**0.00028**
	0.2	**32.08**	**0.9819**	**0.9942**	**0.0231**	**0.0335**	**0.00070**
DnCNN	0.05	35.74	0.9629	0.9858	0.1387	0.0217	0.00028
	0.1	30.67	0.8943	0.9531	0.2376	0.0378	0.00091
	0.2	25.80	0.7388	0.8637	0.3621	0.0626	0.00275
TV	0.05	32.17	0.9752	0.9925	0.0420	0.0360	0.00067
	0.1	27.47	0.9432	0.9793	0.1055	0.0604	0.00198
	0.2	23.40	0.8744	0.9413	0.2431	0.0919	0.00513

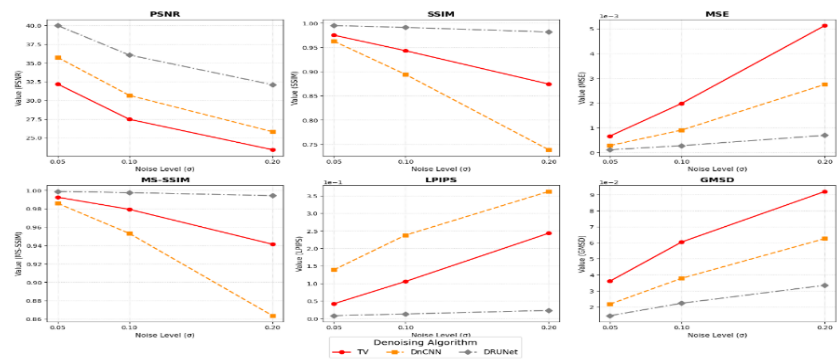

Fig. 5 shows distributions of the metrics for the color images.

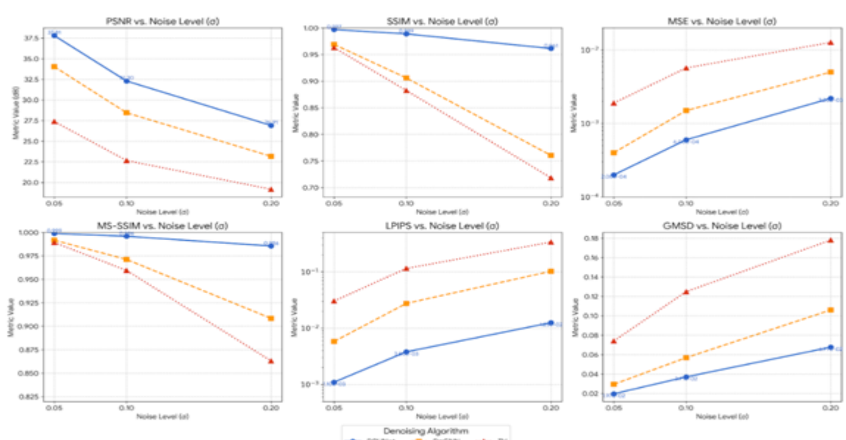

Fig.6 shows distributions of the metrics for the grayscale images.

Table 3: Comparative Analysis of Proposed Work vs. Previous Methods.

Method (Algorithm)	PSNR	SSIM	MS-SSIM	MSE	LPIPS
TV Denoising (Traditional)	24.50	0.9728	0.9791	0.00368	0.0584
DnCNN (Deep Learning)	22.85	0.3911	0.8232	0.00519	0.4137
BM3D (Classical SOTA)	**27.01**	**0.9927**	**0.9938**	**0.00200**	**0.0176**
Proposed Work (DRUNet)	**27.67**	**0.9971**	**0.9978**	**0.00171**	**0.0055**

Comparative Overview and Summative: The relative comparison show that there is an evident performance ranking among the compared models. Compared to TV Denoising and DnCNN, which have issues with structural stability (this leads to prominent blurring and structural loss), the BM3D algorithm is a good classical baseline but does not have the capability of modern architectures to adapt to learning.

Our proposed DRUNet framework is the best solution, which provides the best balance between mathematical accuracy (27.67 dB PSNR) and structural fidelity (0.9971 SSIM). In comparison to the old techniques, our method is effective in removing noise without the loss of fine linguistic details as the LPIPS score is much lower. As a result, the suggested algorithm is the strongest and most perceptually valid restoration and, therefore, the most appropriate option to high-fidelity document image restoration.

8. Discussion

Empirical studies reveal that the modular structure is successful in denoising and deskewing scanned documents and as such improves the overall quality of the digitized inputs. The three algorithms that were analysed in this study are DRUNet, DnCNN, and Total Variation (TV) and the first one performed remarkably better than the rest, at all the noise levels assessed in both greyscale and colour images. The given superiority can be explained by the residual U-Net architecture, which reasonably balances the global contextual understanding and the maintenance of the fine-grained local detail. The experimental results prove that the proposed modular framework suggests a reliable method in improving scanned document quality through denoising and deskewing operations. All the three adopted algorithms namely DRUNet, and DnCNN consistently achieved the best results for all the tested noise levels both for gray and colourful images. This superior performance might be attributed to its retained U-net structure, which effectively combined global contextual information with local information. Quantitative evaluations based on PSNR, SSIM, LPIPS, and GMSD measures indicated that the proposed DRUNet maintained the best subjective and structural fidelity compared to the original images at all the different noise-levels. Although DnCNN demonstrates a good level of noise removal, it also subjects a slight degradation of the text edges in the presence of high levels of noise compared to LQ. The TV algorithm better attenuates the high-frequency noise but at a cost of greater blurring and suppressed texture which negatively influence the text regions. The visual inspections support the empirical results in that DRUNet is suggested to be more faithful to colors and textual clarity than other methods. In addition, the integration of the skew correction preprocessing stage was critical in obtaining reasonable image alignment hence making the restored documents adequate to undergo the OCR processing. The consistency of the suggested system in terms of its performance on color and grayscale images confirms the generality and adaptability of the framework when used on various types of documents and scanners. To sum up, the results support the effectiveness of integration of deep-learning-based denoising algorithms to conventional preprocessing algorithms. This not only makes reconstruction more accurate but also more useful in practical use of the system in digital archiving, automation of administration and maintenance of historical documentations.

9. CONCLUSION

This research focused on improving the quality of scanned document images that are affected by noise and alignment issues. Different denoising approaches were tested, including classical gradient-based methods and deep learning models such as DnCNN. From the obtained results, it was observed that DRUNet provided better performance compared to the other methods, especially in cases with high noise levels. It was able to reduce noise while keeping most of the text edges and important details visible. In addition, the use of preprocessing techniques like Otsu thresholding and deskewing had a clear effect on the final output. These steps helped in improving the appearance of the documents and made the text clearer for further processing. As a result, OCR accuracy was also improved, which is important for document digitization and archiving purposes. Overall, the proposed system can be considered a practical solution for scanned document restoration. It can be applied without complex requirements and gives stable results. In future work, the system may be tested on larger datasets, and further improvements can be made to handle other types of degradation such as blur and low contrast within a single processing framework.

REFERENCES

[1] M. Smith and A. Brown, "Enhancing Document Image Processing: Correcting Skew in Printed Documents Using Deep Learning," *Journal of Image and Graphics*, vol. 13, no. 1, 2025.

[2] R. Zhang, "End-to-End Unsupervised Document Image Blind Denoising," *arXiv preprint arXiv:2101.00000*, 2021.

[3] R. Rotman et al., "A U-Net based pre-processing pipeline for robust OCR with synthetic noisy document datasets," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, 2022, pp. 154–168.

[4] O. Boudraa, W. K. Hidouci, and D. Michelucci, "Using skeleton and Hough transform variant to correct skew in historical documents," *Journal of Mathematics and Computers in Simulation*, vol. 167, pp. 100–114, 2019.

[5] R. Ahmad, S. Naz, and I. Razzak, "Efficient skew detection and correction in scanned document images through clustering of probabilistic hough transforms," *Pattern Recognition Letters*, vol. 152, pp. 93–99, 2021.

[6] W. Kim et al., "A Systematic Review of Deep Learning-Based Image Denoising Methods," *Frontiers in Medical Technology*, vol. 6, Art. 134000, 2024.
[7] H. M. Zangana and F. M. Mustafa, "Hybrid Image Denoising Using Wavelet Transform and Deep Learning," *EAI Endorsed Transactions on AI and Robotics*, vol. 3, pp. 1–10, 2024.
[8] M. S. Tawfik et al., "Comparative Study of Traditional and Deep-Learning Denoising Approaches for Image-Based Petrophysical Characterization," *Frontiers in Water*, vol. 3, Art. 800369, 2022.
[9] H. S. Abdulla, A. S. Shaheen, and N. M. Isaac, "Effectiveness of Image Curvelet Transform Coefficients for Image Denoising," *Al-Rafidain Journal of Computer Science and Mathematics*, vol. 18, no. 2, pp. 1–8, 2024.
[10] H. H. Ali, "Development of Traditional Algorithms and Hybrid Approach for Denoising Color Images," *Al-Rafidain Journal of Computer Science and Mathematics*, vol. 18, no. 2, pp. 36–46, 2011.
[11] A. Supriyono et al., "Advancements in NLP-driven OCR post-processing: A systematic review," *Journal of Digital Document Processing*, vol. 2, no. 1, pp. 45–60, 2024.
[12] G. S. Hukkeri, R. H. Goudar, P. Janagond, and P. S. Patil, "Machine learning in OCR technology: Performance analysis of different OCR methods for slide-to-text conversion," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, 2022.
[13] P. Mohta, "Total variation-based image denoising for edge-preserving smoothing," *International Journal of Computer Vision and Image Processing*, vol. 14, no. 2, pp. 45–58, 2024.
[14] K. Zhang, W. Zuo, and L. Zhang, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 7005–7020, 2021.
[15] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
[16] Y. Zhao et al., "Comprehensive Evaluation of IQA Metrics for Image Restoration Models," *arXiv preprint arXiv:2403.10988*, 2024.
[17] MDPI, "Prospects of Structural Similarity Index for Medical Image Applications," *Sensors*, vol. 22, no. 18, Art. 6890, 2022.
[18] R. Zhang et al., "Learning Perceptual Similarity for Image Restoration Using Deep Feature Representations," *IEEE Access*, vol. 11, pp. 45122–45136, 2023.
[19] M. Safari et al., "MRI Super-Resolution Reconstruction Using Efficient Diffusion Probabilistic Model with Residual Shifting," *Physics in Medicine & Biology*, vol. 70, no. 12, p. 125008, 2025.
[20] K. Loh et al., "A Generalized Quality Assessment Method for Gradient-Based Image Metrics," *IET Image Processing*, vol. 15, no. 12, pp. 2859–2871, 2021.