Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi
Afyon Kocatepe University – Journal of Science and Engineering
https://dergipark.org.tr/tr/pub/akufemubid

Araştırma Makalesi / Research Article
e-ISSN: 2149-3367
DOI: https://doi.org/10.35414/akufemubid.1628066
AKÜ FEMÜBİD **25** (2025) 065101 (1323-1335)
AKU J. Sci. Eng. **25** (2025) 065101 (1323-1335)

A Deep Learning Approach to Document Recovery: High Performance with DenoiseU-Net

Makale Bilgisi / Article Info
Alındı/Received: 28.01.2025
Kabul/Accepted: 14.06.2025
Yayımlandı/Published: 03.12.2025

Belge Kurtarmada Derin Öğrenme Yaklaşımı: DenoiseU-Net ile Yüksek Performans

Salih Can TURAN[1*], Zeki ÇIPLAK[2], Ali SARIKAŞ[3], Kazım YILDIZ[3]

[1] *Marmara University, Institute of Pure and Applied Sciences, Computer Engineering, İstanbul, Türkiye*
[2] *Istanbul Gedik University, Gedik Vacational School, Computer Technologies, İstanbul, Türkiye*
[3] *Marmara University, Faculty of Technology, Computer Engineering, İstanbul, Türkiye*

Abstract

Image denoising, a crucial task in image processing, has consistently faced challenges despite ongoing research efforts. In this study, a dataset was created by extracting 20,000 images from 60 public sources, including scanned or digitized documents. Each image was verified to contain at least one of the following: plain text, image, table, or mathematical expression. Common types of noise, including random black and white pixels, Gaussian blur, gray areas, speckle noise, random directional lines, Poisson noise, and salt-and-pepper noise, were applied to the images. To create the test set, each of the seven types of noise was individually added to 500 images excluded from the dataset, resulting in a balanced test set of 3,500 images. The complete dataset consists of 23,000 images, with a training-to-test ratio of 5:1. Specifically, our proposed DenoiseU-Net model aims to recover noisy scanned documents and performs effectively across various content types, such as tables, images, mathematical equations, and text. Experimental results show that the average precision, recall, and F1-score of DenoiseU-Net on the test set are 99.36%, 99.59%, and 99.48%, respectively. In addition to these evaluation results, the average SSIM and PSNR values, which are commonly used parameters to assess image quality, were obtained as 0.9657 and 40.28 dB, respectively. The primary objective of this study is not to demonstrate superior performance over state-of-the-art (SOTA) methods, but rather to evaluate how deep learning models, such as the proposed DenoiseU-Net, perform on medium-scale or small-scale datasets in practical scenarios.

Keywords: *Document recovery; scanned document; document image denoising; noisy images; deep learning; structural similarity index*

Öz

Görüntü gürültü giderme, görüntü işleme alanında önemli bir görev olmakla birlikte, sürekli araştırmalara rağmen zorluklarla karşılaşmaktadır. Bu araştırmada, 60 farklı kamu kaynağından 20.000 görüntü çıkarılarak bir veri kümesi oluşturulmuştur. Bu görüntülerin her biri, en az bir düz metin, resim, tablo veya matematiksel ifade içerip içermediği açısından kontrol edilmiştir. Görüntülere, rastgele siyah ve beyaz pikseller, Gauss bulanıklığı, gri alanlar, lekelenme gürültüsü, rastgele yönlü çizgiler, Poisson gürültüsü ve tuz-biber gürültüsü gibi yaygın gürültü türleri uygulanmıştır. Test kümesini oluşturmak amacıyla, bu yedi gürültü türü, veri kümesinden hariç tutulan 500 görüntüye tek tek eklenerek 3500 görüntüden oluşan dengeli bir test seti oluşturulmuştur. Veri kümesi, eğitim ve test seti oranı 5:1 olacak şekilde 23.000 görüntüden oluşmaktadır. Özellikle, önerilen DenoiseU-Net modelimiz, gürültülü taranmış belgelerin iyileştirilmesine yönelik olarak geliştirilmiş ve tablolar, resimler, matematiksel denklemler ve metin gibi çeşitli içerik türlerinde başarı göstermektedir. Deneysel sonuçlar, DenoiseU-Net'in test kümesindeki ortalama doğruluk, duyarlılık ve F1-skorunun sırasıyla %99,36, %99,59 ve %99,48 olduğunu göstermektedir. Bu değerlendirme sonuçlarının yanı sıra, görüntülerin kalitesini gösteren yaygın kullanılan parametreler olan SSIM ve PSNR ortalama değerleri sırasıyla 0.9657 ve 40.28 dB olarak elde edilmiştir. Bu çalışmanın ana hedefi, en son teknoloji (SOTA) yöntemleri karşısında üstün performans sergilemek değil, daha çok önerilen DenoiseU-Net gibi derin öğrenme modellerinin, orta ölçekli veya küçük ölçekli veri setlerinde, pratik senaryolarda nasıl performans gösterdiğini değerlendirmektir.

Anahtar Kelimeler: *Belge kurtarma; taranmış belge; belge görüntüsü gürültü giderme; gürültülü görüntüler; derin öğrenme; yapısal benzerlik indeksi*

1. Introduction

Images play a fundamental role in modern information transfer, particularly in digital document processing. With the rapid advancement of technology and the continuous expansion of the digital world, the importance of image processing techniques is increasing day by day. Especially the improvement of image quality plays a critical role in many areas of application (Patel *et al*. 2023, Tahir and Din 2024, Sezer and Altan 2021, Moghadam and Rashidi 2024). However, noise and other distortions in images can seriously affect the quality of images and thus reduce the accuracy and utilization of the information obtained from the image.

Innovations in image processing Technologies (Salvi *et al*. 2021, Rashmi *et al*. 2022) aim to reduce the negative effects of noise on image quality, while helping people to

Sorumlu Yazar / Corresponding Author: Salih Can Turan *e-posta / e-mail: cansalih@marun.edu.tr*

comprehend information more effectively. Noise reduction has emerged as one of the most important steps in image processing. This process aims to suppress the unwanted noise in images and to reveal more clarity and detail (Zhao et al. 2023, Rafiee and Farhang 2023).

The importance of obtaining high quality images is not only aesthetic but also in terms of practical applications. Especially the integration of artificial intelligence and machine learning techniques has revolutionized the field of image processing. Processes such as obtaining high-resolution images from low-resolution images (Dong et al. 2015), enhancing noisy or distorted images (Zhang et al. 2017), digitizing old documents and correcting errors that occur during the scanning process have become more effective and efficient thanks to these technological advances. However, challenges in this area remain. For example, the need for high-quality datasets for training deep learning models, the cost of computational resources, and the adaptation of algorithms to real-world scenarios (Mikołajczyk and Grochowski 2018, Najafabadi et al. 2015) drive researchers toward innovative methods to address these limitations effectively. In this context, advanced technologies such as U-Net (Ronneberger et al. 2015) and attention mechanisms (Oktay et al. 2018) open new horizons in improving image quality and reducing noise, and are considered as key components shaping the future of image processing.

Image denoising to obtain a clean image from a noisy image or to remove the clean area of interest in the image has been actively studied in the field of image processing and still remains a challenging problem (He et al. 2018). Recently, with the development of convolutional neural network (CNN) architecture in deep learning, considerable performance values have been achieved in image denoising (Zhang et al. 2023). Of course, the multi-layered and detailed structure of the architecture also brings with it the need for a high number of parameters and resources. To extract smaller features from the original image, the U-network performs down sampling in the encoding phase. With the decoding process, a higher quality output image is obtained from the input image. The mentioned architecture is frequently used in image denoising (Zhang et al. 2023). With the proliferation of digital devices, image noise reduction has found widespread use in fields as diverse as aviation, beauty, industry and video surveillance. This wide range of applications is a strong motivation to sustain innovations and advances in image processing techniques.

Deep learning has provided groundbreaking results in the field of noise reduction, but the implementation of these methods is challenging due to the complexity of the image denoising process. In response to these challenges, innovative structures such as U-Net and attention mechanisms have significantly improved image processing.

U-Net has achieved particularly successful results in medical image segmentation, providing in-depth learning and information integration in a wide context. This model has a symmetric structure and is characterized by its "U" shaped design. The main motivation of U-Net is its ability to perform detailed segmentations in high-resolution images even from a small number of training samples. The model is characterized by up-sampling paths and jump links, which allow combining high-level features from deep layers of the network with low-level features closer to the input layer. This feature makes U-Net very effective in processes such as reducing noise from the image and improving its quality. Because the model can effectively capture features at different scales of the image and thus produce more detailed and accurate results. Attention mechanisms determine the parts of the image that the model should focus on, enabling dynamic adjustment of the receptive field and thus more accurate feature extraction.

Our main motivation has been to develop a deep learning model that fulfills a general need for a solution to provide a wide range of use cases by recovering different content such as text, images, tables, etc. from scanned or digitized documents. The primary contributions of this study are clearly outlined as follows:

- The DenoiseU-Net deep learning model is specifically developed for the restoration of scanned documents that contain a wide variety of noises.
- The implemented deep learning model is adept at recovering scanned documents regardless of their content such as text, tables or images.
- Innovative approaches have been introduced for pre-processing the inputs, increasing their suitability for deep learning models, especially for the restoration of extremely large documents.

The paper's organization consists of the following sections: Section 2 situates the study by reviewing similar research in the literature and provides a summary of the findings, methods and limitations of previous studies. Section 3 elaborates on the customized data collection strategies and preprocessing methods suitable for the research objectives.

This section also describes the proposed deep learning model. Section 4 presents the findings of the research and

discusses the implications and significance of these findings. Section 5 summarizes the key findings of the study and offers future research recommendations.

2. Related Works

In their study (Dong et al. 2018), Chao Dong et al. proposed a deep learning-based method to obtain Super Resolution (SR) from a single image. They developed a CNN model that learns direct end-to-end mapping between low-resolution images and high-resolution images. This model is able to reformulate traditional sparse coding-based SR methods as deep CNNs. The method, called SRCNN, has a simple architecture. The results show that the proposed SRCNN model significantly outperforms the best existing SR methods. Using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics, the SRCNN model achieved the highest average PSNR values compared to other methods on Set5 and Set14 datasets and outperformed the bicubic interpolation method even early in the learning process. For example, for upscaling factor 3, the average PSNR value obtained with SRCNN in the Set5 dataset was found to be 32.39 dB.

Pascal Vincent et al. introduced noise removal autocoders in their work (Vincent et al. 2019). The authors proposed to use this new training principle to initialize deep architectures that learn to reconstruct clean and repaired inputs from noise-added inputs. It is shown that deep learning models can be significantly improved by the use of noise removal autocoders. In experiments using the MNIST dataset, lower error rates were obtained thanks to this method. Experiments confirmed that this approach provides better classification performance than existing methods on various tasks. Rudin et al. (Rudin et al. 1992) presented a restricted optimization type algorithm for image noise removal based on total variation minimization. The constraints on the statistics of the noise are implemented using Lagrange multipliers and the solution is obtained by gradient-projection. The method involves solving time-varying partial differential equations to reach a steady state that converges to a denoised image. It is also a non-intrusive method that achieves good results for very noisy images and provides sharp edges in the images. The method is described as a two-step procedure. In the first step, each level set of the image is moved in the direction normal to itself at a speed equal to the curvature of the level set divided by the magnitude of the image gradient. In the second step, the resulting image is projected back onto the constraint set.

In their study on bilateral filtering, C. Tomasi and R. Manduchi (Tomasi and Manduchi 1998) tested an edge-preserving image smoothing method for grey and color images. Bilateral filtering preserves edges while smoothing colors in both grey level and color images in a way suitable for human perception. Experiments showing the effects of this filtering method on various grey and color images are presented in the study. The results show that double-sided filtering effectively reduces noise in the image while preserving edges and prevents color distortion while smoothing color transitions in color images.

Antoni Buades and Jean-Michel Morel (Buades et al. 2005) proposed the measurement of 'method noise' to evaluate and compare the performance of noise reduction methods from digital images. Method noise is expressed in terms of perceptial quality and mean square error (MSE) and is defined as the difference between the original (always with some noise) image and its denoised (noise-free) version. Visual quality was assessed by the human eye judging whether the denoising method improved the image or not. The MSE was calculated as the Euclidean difference between the restored and the real images, which objectively indicates how close the estimate is to the original image. The results of the study prove that the Non-Local Means (NL-means) algorithm outperforms conventional local smoothing filters, especially in low and high noise images.

In their study (Huang et al. 2025), Jia-Bin Huang et al. aimed to obtain super resolution from a single image. The goal of the study is to construct an extended internal search space that takes into account the geometric variations of iterative patterns within the image. This search space is structured using additional affine transformations to model the planar perspective geometry and local shape variations. In the proposed method, datasets commonly used in computer vision studies (Urban and BSD) are tested. 'Urban' usually refers to datasets containing images of urban scenes (such as buildings, roads, vehicles), while Berkeley Segmentation Dataset, BSD, is another dataset containing images of natural scenes. PSNR and SSIM metrics were used in the evaluation of the study. According to the results of the study, significantly better results were obtained in urban scenes (Urban dataset), while a performance comparable to the best available algorithms was achieved in natural scenes. The best performances were 29.05 dB for 2x scale and 24.67 dB for 4x scale in PSNR metric; 0.8980 for 2x scale and 0.7314 for 4x scale in SSIM metric.

Jiwon Kim et al. (Kim et al. 2016) presented a high-accuracy SR image retrieval method using very deep convolutional networks, which goals to reconstruct a

high-resolution image using only one low-resolution image. This method is inspired by VGG-Net and has a very deep network with 20 weight layers. During training, it is proposed to accelerate the training process by learning only residuals and adjustable gradient clipping. The proposed method provided superiority in both accuracy and visual improvements compared to existing methods. Especially in terms of PSNR and SSIM metrics, it outperformed the existing methods in various test sets. In the Set5 dataset, the PSNR value for the x2 scale factor gave the best result with 37.53 dB. For the same data set and the same scale factor, the SSIM metric also gave the best result with a value of 0.9587.

Yan Jiang et al. (Jiang et al. 2017) carried out a study on the restoration technology of scan images of old documents. As a method, a high-definition scanner was used. With this scanner, documents were scanned page by page to obtain HD images, followed by various processing techniques such as removing black spots, deepening handwriting and reducing noise. The paper shows that this method is simple and feasible, preserving the original version of the old files and facilitating open and efficient human access to these documents. Specific metrics indicating the accuracy of the results are not explicitly stated, but the method was effective on almost 10,000 images, making electronic documents easily accessible to humans.

Kathrin Berkner (Berkner 2001) carried out a study on the reduction of blurring or noise in documents after scanning. In the study, Wavelet-based noise reduction and sharpening methods were tested using Besov spaces and taking advantage of the advantages of processing in the Wavelet domain instead of the Fourier domain. The proposed method is designed for the enhancement of composite documents containing various types of images such as text, halftone patterns, background and photographic images. Experimental results show that the proposed Wavelet-based enhancement system is qualitatively superior to conventional Fourier-based techniques. In particular, it is found to effectively reduce background noise and noise due to halftone patterns while improving text sharpness.

Darshil Mehta et al. (Mehta et al. 2022) used the U-Net architecture and various image processing techniques to clean noisy MRI scans. By fine-tuning the U-Net architecture with two encoder-decoder pairs on a data set injected with synthetic Gaussian noise, the authors improved the PSNR from 11.90 to 30.96. It is also experimentally proven that the proposed noise removal method improves the brain tumor prediction accuracy by about 23%. These results show that the U-Net based noise reduction approach is highly effective in cleaning MRI scans. This improvement is important for both visual quality and medical image analysis and disease diagnosis.

Mohit Kulkarni et al. (Kulkarni et al. 2020) focused on the development of an algorithm that aims to remove text from old documents and make them readable. In the study, they converted RGB color images to grayscale. Methods such as histogram, median filter, noise subtraction and image segmentation were applied to reduce image noise. The median filter was found to effectively reduce the salt-and-pepper noise and other types of noise. As a result, noises such as coffee stains were reduced and the text was converted into a cleaner and more readable format.

Pranjal Jadhav et al. (Jadhav et al.2022) used a Pix2Pix Generative Adversarial Network (GAN) model developed with ResNet to reduce the noise of electronic document images corrupted during the scanning process. The developed model uses a generator network where ResNet6 replaces the U-Net architecture and uses patchGAN as discriminator. For training, a noisy document dataset generated by adding synthetic noises was used. For model evaluation, quantitative SSIM and PSNR metrics and qualitative Optical Character Recognition (OCR) tests were performed. The average SSIM value obtained was 0.9108 and the average PSNR value was 31.8 dB.

In their study (Hsu et al. 2022), Enshuo Hsu et al. performed several tests on information extraction from scanned documents in electronic health records (EHR). The main goal of the study was to extract Apnea Hypopnea Index (AHI) and Oxygen Saturation (SaO2), two main indicators for sleep apnea, from 955 scanned sleep study reports. We experimented with known image preprocessing methods such as grayscale conversion, stretching, erosion and contrast, Tesseract OCR, various machine learning and deep learning models. Additionally, they assessed two configurations of deep learning architectures: one with structured input that offers details on document layout, and the other without. The results of the study showed that the document accuracy of the proposed method was 94.76% for AHI and 91.61% for SaO2.

Shubham Paliwal et al. (Paliwal et al. 2019) proposed a deep learning model, "TableNet" that can be used for table detection and table content extraction from scanned document images. It is designed to perform the tasks of table detection and table content recognition simultaneously. A pre-trained network of VGG-19 is utilized for feature extraction. It includes two separate

decoders for segmentation of table and column regions. This dual-tasking structure has enabled TableNet to achieve effective results in table detection and table data extraction. The model was tested on the ICDAR 2013 and Marmot Table datasets and achieved high metric values. In table detection, recall rate was 96.28%, Precision rate was 96.97% and F1-Score was 96.62%. In table structure recognition and table content extraction tasks, the Recall rate was 90.01%, Precision rate was 93.07% and F1-Score was 91.51%.

In their research (Srinivasa et al. 2019), Srinivasa et al. addressed the problem of noise reduction in the process of digitizing information from ancient texts, stone tablets and similar documents. The aim of the project is to use neural networks and image processing techniques to eliminate noise from photographs of these papers. The methods used include edge detection, thresholding, filtering, morphological operations and CNN. The results of the study show that the RMSE value obtained with the applied image processing techniques is 0.0685, while the RMSE value obtained using CNN is 0.0152. These results show that the use of CNN significantly improves the accuracy, resulting in almost five times more effective noise reduction.

Shubham Kumar Gupta et al. (Gupta et al. 2023) conducted a study on noise reduction in acoustic microscopy images. The method used in the study is the block matching and four-dimensional filtering (Block-Macthing 4D - BM4D) technique. BM4D includes hard thresholding and Wiener filtering stages with transform domain filtering technique on noisy images. Comparative experiments with conventional filters (Gaussian, median, Wiener filters) have shown that the BM4D filter is more effective in reducing the noise of acoustic images. SSIM and PSNR metrics were used for qualitative and quantitative analysis. The results of the study show that BM4D performs the best in terms of SSIM and PSNR. For 0.24Vpp, the PSNR value was 31.88 dB and the SSIM value was 0.87. For 0.25Vpp, the PSNR value reached 39.78 dB and the SSIM value reached 0.97. These results show that the BM4D filter can effectively enhance noisy acoustic images and is especially suitable for situations with low signal-to-noise ratio.

Kreuzer (Kreuzer and Munz 2023) used a Swin Transformer-based U-Net structure to eliminate artifacts (errors) during the scanning process. The goal of the work is to improve text extraction quality by reducing errors in the OCR process. The paper presents a method to learn features at certain levels more selectively using multi-headed cross-attention skip connections. The model is analyzed for its performance in removing typical scan document artifacts. The results show an improvement in text extraction quality on synthetic data with an error rate reduction of up to 53.9%. For pixelation and compression errors, 37.1% and 36.0% improvement are achieved, while 53.9% improvement is achieved for noisy images.

Zulkarnain et al. (Zulkarnain et al. 2022) focused on extracting the content of tables with unclear boundaries in image documents. In this context, the deep learning model Mask R-CNN-FPN (a combination of Region-based Convolutional Neural Network and Feature Pyramid Network) and data augmentation techniques were used. The data augmentation technique used is fine-tuning with the CutMask augmentation method. The UNLV dataset was used for model building and testing, starting with 427 samples in total and increasing to 854 samples after data augmentation. The test results show data augmentation contributes to better accuracy in deep learning models in detecting tables with unclear boundaries. However, the expected results were not achieved in table structure recognition. Especially for the method used in data augmentation, the accuracy value was 73.5%, precision 86.3%, recall 93.2% and F1 score 89.6%.

Hu et al. (Hu et al. 2021) discussed deep learning approaches for determining the optimal resolution of scanned text documents. In the study, a novel CNN-based model is proposed to estimate the minimum reusable resolution (MRR) of documents or document regions. The methods include a compression framework that separates document images into symbols, rasters and vectors and estimates optimal scanning settings for these regions; a page segmentation algorithm that applies specific algorithms to specific regions; and a CNN-based model that estimates MRR. In the tests, MobileNetV2, MultiscaleNet and MultiscaleNet-IRB models were compared and it was reported that MobileNetV2 performed the best with 93.11% test accuracy, but MultiscaleNet and MultiscaleNet-IRB performed similarly with much fewer parameters and computational requirements. These models provide lightweight and efficient solutions suitable for real-time applications on ARM-based CPUs.

Schreiber et al. (Schreiber et al. 2017) proposed a new end-to-end system called DeepDeSRT for the detection and structure recognition of tables in documents. The work has two main goals. The first one is to detect tables in document images, and the second one is to recognize the structure of the tables describing their row-column and cell positions. For both of these objectives, innovative deep learning-based approaches are proposed which is

completely data-driven without the need for heuristics or PDF metadata. Evaluations on the ICDAR 2013 table competition dataset show that it is superior to existing methods with an F1 score of 96.77% for table detection and 91.44% for structure recognition.

Yin and friends (Yin et al. 2023) proposed the SWA-U2former, an efficient noise removal network with a nested double U-shaped architecture to remove complicated noise from images. The attention mechanism and image convolution are combined in the SWA-RSU (Shifted Window Attention Residual U-blocks) module. In the SWA-RSU block, multiple SWA-Transformer blocks are stacked in a U-shaped structure. This structure improved the efficiency of noise removal and restoration and preserved image details better. The attention mechanism helped CNN to better cope with image noise and distortion. The tests showed that SWA-U2former achieved PSNR values of 40.06 dB and 40.10 dB on the SIDD and DND datasets, respectively. The SSIM metric values were also measured at 0.955 and 0.960 respectively.

Mange et al. (Mange et al. 2023) studied a CNN model that uses feature selection, guided by an attention mechanism, to provide an improvement for removing noise from images. In the method used in the study, there are two CNN models and each model has an attention mechanism attached to its output. This attention mechanism allows the model to select important features and combine them dynamically. The results show that this method, called CNWATT2, surpasses existing noise removal models (DnCNN, FFDNet, IRCNN and BRDNET) in terms of both objective (using PSNR and SSIM metrics) and subjective quality. It is found to be capable of effectively removing Gaussian noise, especially in color and grayscale images. For noise level σ=75, the CNWATT2 model performed the best, achieving PSNR values of 34.7 on the CBSD68 dataset, 35.23 on the Kodak24 dataset and 37.3 on the CT images. In terms of SSIM values, for σ=15 noise level, the CNWATT2 model achieved high values of 0.9617 in CBSD68 dataset, 0.9625 in Kodak24 dataset and 0.9666 in CT images.

Liu et al. (Liu et al. 2023) introduced RA-UNet, a novel network model for image noise removal. RA-UNet is inspired by the classical U-Net architecture, integrating multiple Residual Convolutional Blocks and attention mechanism that can adapt to different scales. Based on the concept of Noise2Noise, this model trains a neural network using noisy image pairs. Compared to BM3D, DnCNN, ADNet, RED30 and UNet, the noise removal performance of RA-UNet is significantly better in terms of PSNR and SSIM for some datasets. At σ=25 noise level, Gaussian gave the best result with a SSIM value of 0.9505 for the Set14 dataset compared to the aforementioned models. For σ = 50, it was found to give the best PSNR value with 31.13 in the B100 dataset. At a noise level of p = 0.15, it was observed that it reached 0.9601 SSIM value for Urban100 dataset and 0.9563 SSIM value for Set14 dataset.

3. Material and Methods

3.1 Data Generation and Preprocessing

To create the dataset, we first used scanned or digitized documents from 60 public sources, some of which we owned. Then, from these documents, image areas with a size of 256x256 pixels were randomly selected to create the dataset. A total of 20,000 images were extracted and it was ensured that the selected images contained at least one of the following: text, images, mathematical expressions, tables, etc. Seven common synthetic noise types (black-white pixels, Gaussian blur, gray patches, speckle noise, random directional lines, Poisson noise, salt-and-pepper noise) were systematically introduced to the images. The parameters and brief descriptions of these noises are summarized in Table 1.

Table 1. Synthetic noise models, their parameters, and brief descriptions used in document image degradation

Noise Type	Parameter(s)	Brief Description
Random Black & White	blackProb=0.12, whiteProb=0.12	Generates random black or white pixels, adding a grainy appearance and reducing contrast.
Gaussian Blur	blur_strength=0.5	Blurs the image, which decreases edge sharpness and text clarity.
Gray Areas	noise_prob=0.4	Inserts random gray patches into the image, impairing overall readability.
Speckle Noise	speckle_intensity=0.4	Introduces dotted or speckled artifacts that degrade text visibility.
Random Lines	line_num=500, line_thickness=1, line_color=192	Draws random thin lines across the document, potentially obscuring characters or tables.
Poisson Noise	Poisson distribution (no explicit parameter)	Simulates low-light or quantum-related noise, creating random variations in pixel intensity.
Salt & Pepper Noise	salt_prob=0.1, pepper_prob=0.1	Disrupts legibility by adding white or black specks throughout the image.

Out of these, 500 images were excluded to form the source images of the test set. Seven different noises were added to these images to increase the number of test set images to 3,500 to measure the recovering performance of the deep learning model. As a result, the dataset consists of 23000 images. The ratio of the training and test set is approximately 5:1. These steps resulted in an enriched dataset with processed images that provided an environment of varying difficulty levels for training deep learning models.

The whole process of noise removal from documents includes the generation of custom dataset, data preprocessing, architectural design and training of the deep learning model, and evaluation of the test dataset. Hierarchically, these processes are shown in the Fig. 1.

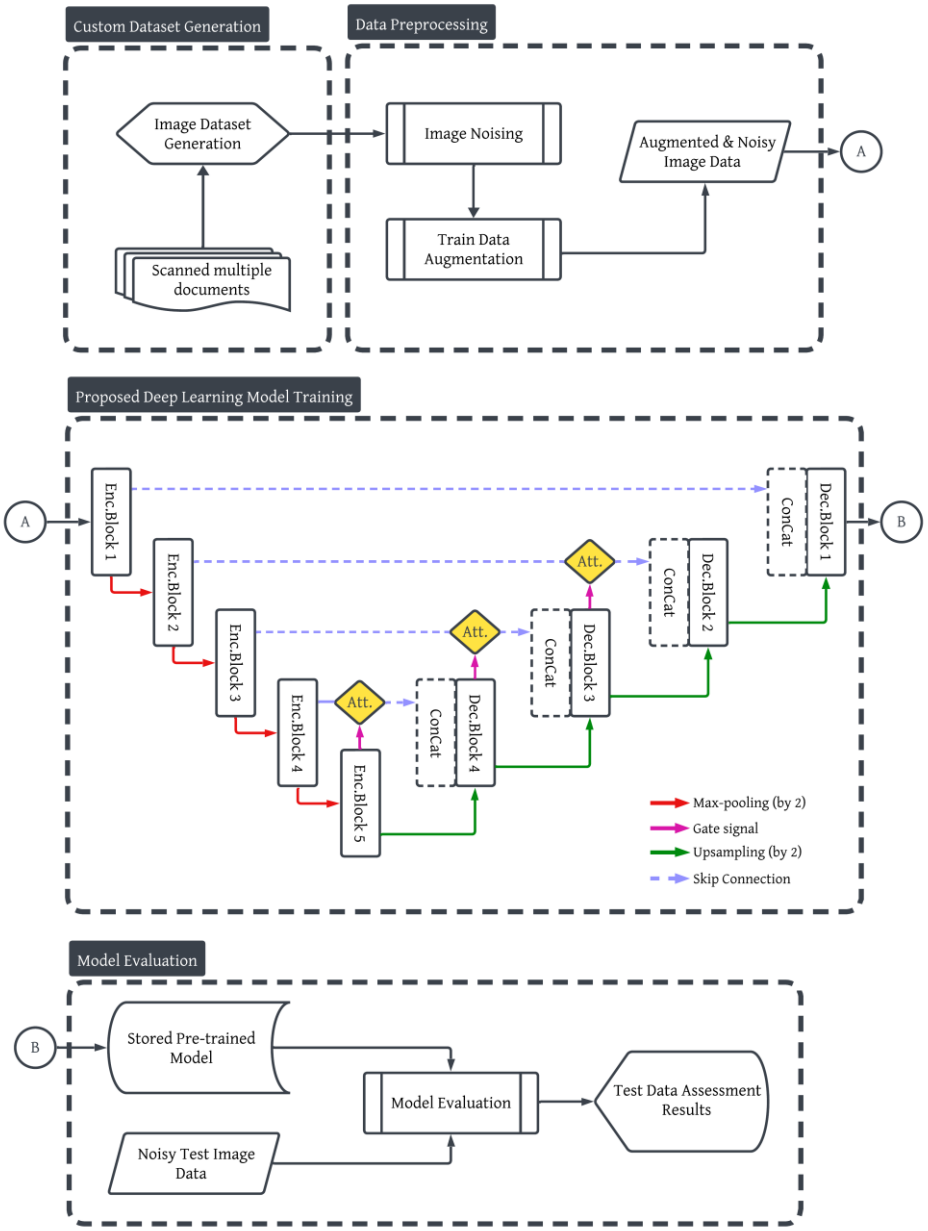

Figure 1. General framework for document denoising

3.2 Proposed Deep Learning Model: DenoiseU-Net

The PyTorch model proposed in this paper describes a deep learning model called “DenoiseU-Net”. The network is designed to extract clean images from noisy images. The model is based on the U-Net architecture and includes attention mechanism blocks [13, 14].

The blocks in the DenoiseU-Net model architecture can be briefly described as follows:

- Double convolution layer: It is a two-dimensional convolution layer. It is used to extract feature maps from the input image. Kernel size convolves over a given window size and learns different features of the image.
- Rectified Linear Unit: It is a non-linear activation function. It converts negative inputs to zero and positive inputs unchanged. This allows the network to learn faster and reduces overfitting.

- Batch Normalization: It speeds up the training process and increases the stability of the network by performing normalization in the post-convolution layer.
- Sigmoid: It is an activation function with boundaries between 0 and 1. In this model, it is used to normalize the output and compress the pixel values between 0 and 1.
- Average Pooling: Implements the average pooling process. This is used to reduce the size and use fewer parameters. The output minimizes a feature map.
- Attention Block: This block computes attention weights by processing input and feature maps. It learns the relationship between input and feature maps and then computes attention weights.
- Wg: This part forms a network that carefully inspects the signal coming from the input. It consists of two parallel convolution layers and a normalization layer. First, a Conv2d layer is used to extract the features of the input signal. Then, ReLU is used as the activation layer and BatchNorm2d as the normalization layer.
- Wx: This section represents a mesh that extracts features from the feature map. Again, it consists of Conv2d, ReLU and BatchNorm2d layers.
- psi: This part is used to generate the attention score. First, information is integrated across the feature map with a Conv2d layer. Then, the attention score is calculated using a sigmoid activation function.

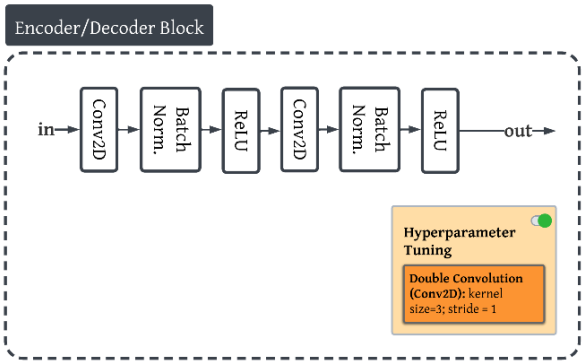

Figure 2. Encoder and decoder blocks

Figure 2 shows the encoder and decoder blocks embedded in the proposed deep learning model. These perform the 'double convolution layer (Conv2D)', 'batch normalization' and 'ReLU activation function' twice.

Fig. 3 shows the attention blocks in the proposed model. The DenoiseU-Net model was implemented using PyTorch and trained for 70 epochs with the Adam optimizer, a learning rate of 0.001, a batch size of 32, and Mean Squared Error (MSE) as the loss function.

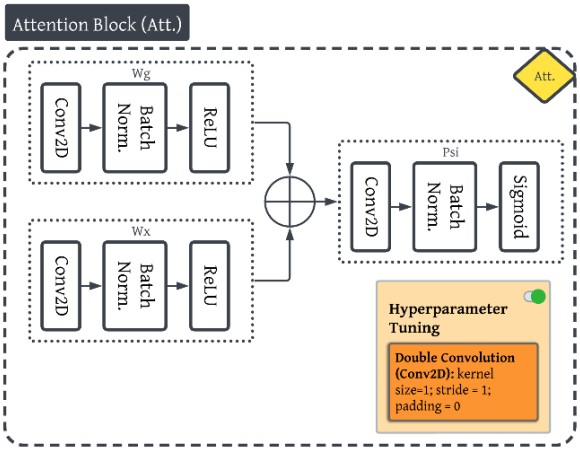

Figure 3. Attention mechanism [14]

3.3 Evaluation Metrics

3.3.1 Peak Signal-to-Noise Ratio

The PSNR is the relationship between an image's maximum potential power and the distorting noise power that reduces the image's representational quality (Alshathri et al. 2022). Equation 1 for the definition of PSNR:

$$PSNR = 20log_{10}\left[\frac{M-1}{RMSE}\right] \quad (1)$$

Here, M is the maximum possible number of intensity levels in an image and the root mean square error. The formula for RMSE is computed as equation 2:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\|y_i - \hat{y}_i\|^2}{N}} \quad (2)$$

Where, N is the total number of pixels, y_i is the intensity of the i-th pixel in the original image, and $\hat{y}_i$ is the intensity of the corresponding pixel in the recovered or denoised image.

3.3.2 Structural Similarity Index

The degree to which the overall structure of the image is conserved is measured by the structural similarity index, which represents the "perceptual quality" of the image (Alshatri et al. 2022, Wang et al. 2004). The similarity index between two grayscale images p and q is computed as equation 3:

$$SSIM = \frac{(2\mu_p\mu_q + c_1)(2\sigma_{pq} + c_2)}{(\mu_p{}^2 + \mu_q{}^2 + c_1)(\sigma_p{}^2 + \sigma_q{}^2 + c_2)} \quad (3)$$

Here, μ_p and μ_q are the average intensities, $\sigma_p{}^2$ and $\sigma_q{}^2$ are the variances, and σ_{pq} is the covariance of the two

images. The constants c_1 and c_2 are introduced to prevent instability caused by small denominators.

SSIM index satisfies the following conditions:

- Symmetry: $SSIM(p,q) = SSIM(q,p)$;
- Boundedness: $SSIM(p,q) \leq 1$;
- Unique maximum: $SSIM(p,q) = 1$ if and only if p=q.

An SSIM score closer to 1 implies a stronger resemblance in structure between images.

3.3.3 Precision, Recall, and F1-Score in Image Comparison

Although precision, recall, and F1-score are traditionally employed in classification tasks, these metrics can be adapted for evaluating grayscale image quality by performing pixel-wise comparisons after binarization. In this study, both the original and denoised grayscale images were converted into binary form using a fixed global thresholding method with a threshold value of 128. This value was chosen to separate foreground (object) and background regions, enabling a meaningful comparison of structural elements in the images. Following binarization, pixel-level classification is performed by categorizing each pixel as true positive, false positive, or false negative, thereby allowing the computation of precision, recall, and F1-score as indicators of structural fidelity. The binarized images are then compared pixel-by-pixel:

- True Positive (TP): A pixel correctly predicted as foreground.
- False Positive (FP): A background pixel incorrectly predicted as foreground.
- False Negative (FN): A foreground pixel incorrectly predicted as background.

From this binary classification of pixels, the following metrics are computed:

Precision (*Pr*) measures the proportion of predicted foreground pixels that are actually foreground in equation 4:

$$Pr = \frac{TP}{TP+FP} \tag{4}$$

Recall (*Re*) evaluates the proportion of true foreground pixels that are correctly detected in equation 5:

$$Re = \frac{TP}{TP+FN} \tag{5}$$

F1-Score is the harmonic mean of precision and recall, providing a balanced measure is shown in equation 6:

$$\text{F1-Score} = \frac{2\times \text{Pr}\times \text{Re}}{\text{Pr}+\text{Re}} \tag{6}$$

These metrics allow for a more nuanced assessment of how well the denoised or reconstructed image preserves object boundaries and important structural information when compared to the original.

4. Results and Discussion

In this section, we present the original, degraded (noisy), and recovered versions of eight representative images selected from the 3,500-image test set, which comprises 500 images for each synthetic noise type. Table 2 provides a quantitative evaluation of the proposed DenoiseU-Net model's performance on these individual samples, based on structural similarity (SSIM), peak signal-to-noise ratio (PSNR), and detection metrics.

Table 2. Quantitative Evaluation of Image Quality and Model Performance on Test Samples

Test Image	SSIM	PSNR (dB)	Precision (%)	Recall (%)	F1-score (%)
#1	0.9054	31.53	99.31	99.30	99.31
#2	0.8578	34.32	98.16	99.05	98.61
#3	0.9298	35.50	99.24	99.60	99.42
#4	0.8756	33.89	97.16	98.99	98.07
#5	0.9233	35.83	99.12	99.61	99.36
#6	0.7844	32.35	98.36	97.63	97.99
#7	0.9518	34.30	98.64	98.65	98.65
#8	0.9604	38.09	99.30	99.55	99.43

The test images presented in Figure 4 were selected to represent examples that differ solely in terms of content type (e.g., tables, mathematical expressions, plain text, etc.). The aim of this selection is to evaluate the visual performance of the proposed method in the presence of varying content types, independent of the noise type present in the images. In this way, the generalizability of our method across different document types and its sensitivity to content structure are more clearly demonstrated.

The noise type-based and overall performance results of the proposed DenoiseU-Net model on 3,500 test images are given in Table 3. Overall results demonstrate that our method is consistent and effective in document restoration. The results obtained from each of the test images are very close to each other, as evidenced by the very small standard deviation values.

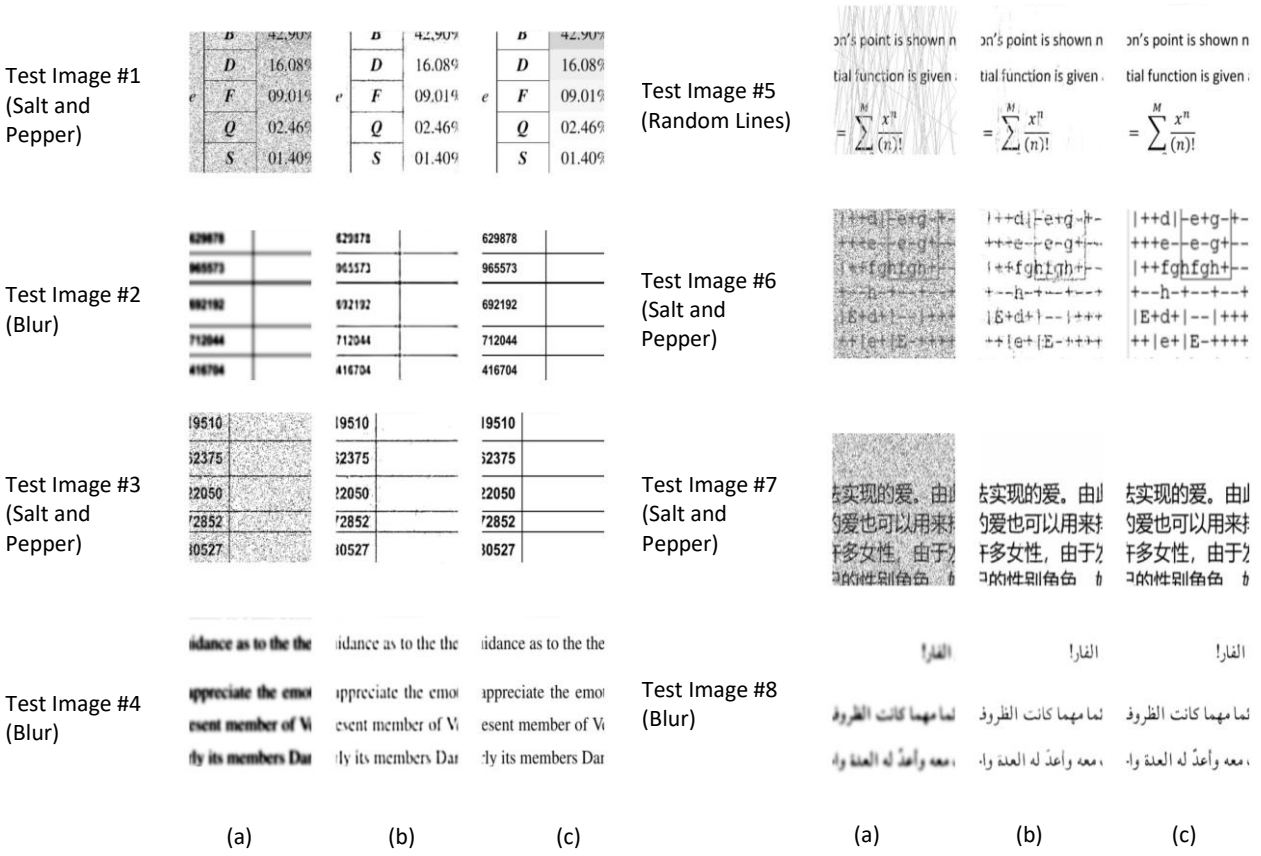

Figure 4. Results of some test images, (a) degraded, (b) recovered and (c) original

Table 3. The noise type-based and overall performance results of the proposed DenoiseU-Net model

Noise Type		SSIM	PSNR (dB)	Precision	Recall	F1-Score
Black and White Pixels	*Mean*	.9654	41.22	.9943	.9954	.9948
	SD	.0403	5.53	.0093	.0054	.0068
Speckle Intensity	*Mean*	.9748	40.54	.9947	.9974	.9960
	SD	.0390	5.53	.0078	.0038	.0056
Gray Areas	*Mean*	.9608	40.51	.9924	.9953	.9938
	SD	.0457	5.71	.0086	.0055	.0068
Random Lines	*Mean*	.9449	41.51	.9957	.9976	.9967
	SD	.0782	5.48	.0085	.0045	.0057
Blur	*Mean*	.9566	39.88	.9887	.9959	.9922
	SD	.0486	5.59	.0134	.0044	.0086
Poisson	*Mean*	.9862	41.49	.9980	.9981	.9980
	SD	.0268	5.29	.0062	.0040	.0045
Salt and Pepper	*Mean*	.9410	41.09	.9950	.9958	.9954
	SD	.0767	5.42	.0067	.0048	.0054
Overall	*Mean*	.9657	40.28	.9936	.9959	.9948
	SD	.0416	5.20	.0080	.0049	.0061

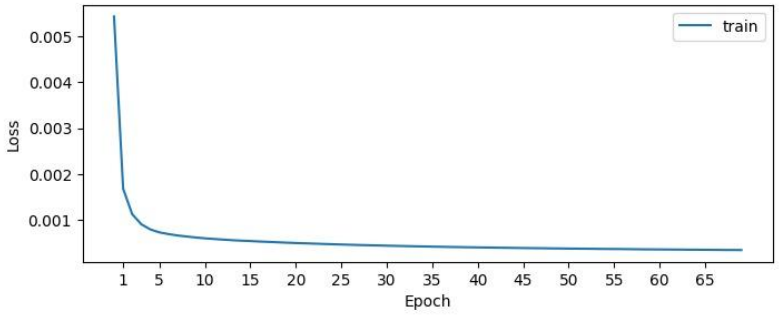

Figure 5. Loss curve for the training set

Figure 5 shows the curve of the loss values of the training set over 70 epochs in the training process. As can be seen from the chart, the loss value during the training process is well below 0.001.

Figure 6 illustrates the progression of SSIM and PSNR values computed on a dedicated subset of the training data, which was reserved solely for monitoring purposes and does not constitute an independent validation set. The upward trend observed in both metrics throughout the training process indicates a continuous improvement in image reconstruction quality, suggesting that the model is effectively learning to reduce noise without overfitting to the training data.

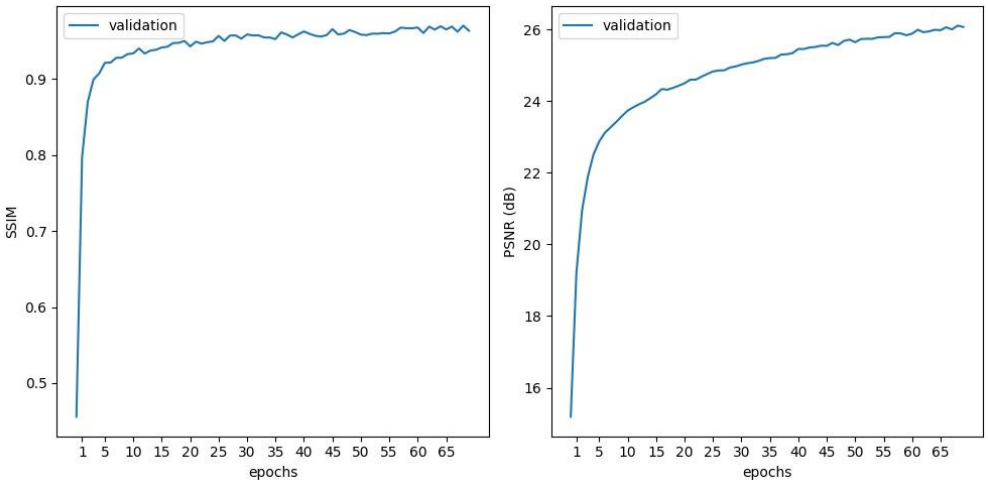

Figure 6. Evolution of SSIM and PSNR metrics per epoch on the monitoring subset

5. Conclusion

Recent advances in CNN architectures have significantly promoted the effectiveness of image denoising procedures. Utilizing data from a variety of sources, a dataset of 60 documents was rigorously constructed to stimulate the development of a versatile deep learning framework tailored for the restoration of content from scanned documents. This dataset served as a touchstone for the inception of the DenoiseU-Net model, which was thoroughly designed to excel at the complex task of recovering various content modalities common in scanned documents. The experimental evaluation of the DenoiseU-Net model has demonstrated its achievement in recovering tables, images, equations and textual elements from noisy document scans, as evidenced by its commendable precision, recall and F1 score on the test set. Furthermore, the model demonstrated robust performance through high SSIM and PSNR metrics. Further research is planned to increase the size of the dataset and evaluating the model on real-world document archives to further improve its generalizability and robustness.

Declaration of Ethical Standards
The authors declare that they comply with all ethical standards.

Credit Authorship Contribution Statement
Author-1: Methodology / Study design, Software, Validation, Formal analysis, Investigation
Author-2: Software, Validation, Formal analysis, Investigation
Author-3: Writing – original draft, Writing – review and editing, Supervision
Author-4: Investigation, Writing – review and editing, Supervision

Declaration of Competing Interest
The authors have no conflicts of interest to declare regarding the content of this article.

Data Availability Statement
All data generated or analyzed during this study are included in this published article.

6. References

Alshathri, S. I., Vincent, D. J., & Hari, V. S. (2022). Denoising Letter Images from Scanned Invoices Using Stacked Autoencoders. *Computers, Materials & Continua*, **71(1)**, 1371-1386. https://doi.org/10.32604/cmc.2022.022458

Berkner, K. (2001). *Enhancement of scanned documents in Besov spaces using wavelet domain representations*. Document Recognition and Retrieval IX. San Jose, CA, USA, 143–154.

Buades, A., Coll, B., & Morel, J. M. (2005). *A non-local algorithm for image denoising*. IEEE computer society conference on computer vision and pattern recognition San Diego, CA, USA, 60-65.

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. https://doi.org/10.1109/TPAMI.2015.2439281

Dong, C., Loy, C. C., He, K., & Tang, X. (2014). *Learning a deep convolutional network for image super-resolution*. 13th European Conference on Computer Vision (ECCV 2014). Zurich, Switzerland, 184-199.

Gupta, S. K., Pal, R., Ahmad, A., Melandsø, F., & Habib, A. (2023). Image denoising in acoustic microscopy using block-matching and 4D filter. *Scientific Reports*, 13, 13212. https://doi.org/10.1038/s41598-023-40301-7

He, W., Zhang, H., Shen, H., & Zhang, L. (2018). Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **11(3)**, 713-729. https://doi.org/10.1109/JSTARS.2018.2795093

Hu, L., Hu, Z., Bauer, P., Harris, T. J., & Allebach, J. P. (2021). Deep learning approaches to determining optimal resolution for scanned text documents. *Electronic Imaging*, **33**, 1-8. https://doi.org/10.2352/ISSN.2470-1173.2021.16.COLOR-243

Huang, J. B., Singh, A., & Ahuja, N. (2015). *Single image super-resolution from transformed self-exemplars*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA, 5197-5206.

Hsu, E., Malagaris, I., Kuo, Y. F., Sultana, R., & Roberts, K. (2022). Deep learning-based NLP data pipeline for EHR-scanned document information extraction. *JAMIA open*, **5(2)**, 1-12. https://doi.org/10.1093/jamiaopen/ooac045

Jadhav, P., Sawal, M., Zagade, A., Kamble, P., & Deshpande, P. (2022) . *Pix2pix generative adversarial network with resnet for document image denoising*. 4th International Conference on Inventive Research in Computing Applications (ICIRCA). Coimbatore, India, 1489-1494.

Jiang, Y., Xia, Y., Feng, Z., & Li, J. (2017, July). *Research on the recovery technology of scanned image of obsolete document*. 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS), Dalian, China, 2017, 61-65.

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

Kreuzer, D., & Munz, M. (2023). Transformer-Based UNet with Multi-Headed Cross-Attention Skip Connections to Eliminate Artifacts in Scanned Documents. *arXiv preprint arXiv:2306.02815*. https://doi.org/10.48550/arXiv.2306.02815

Kulkarni, M., Kakad, S., Mehra, R., & Mehta, B. (2020). Denoising documents using image processing for digital restoration. Proceedings of ICMLIP Machine Learning and Information Processing. Singapore, 287-295.

Liu, W., Li, Y., & Huang, D. (2023). RA-UNet: an improved network model for image denoising. *The Visual Computer*, **40(6)**, 4319-4335. https://doi.org/10.1007/s00371-023-03084-6

Mange, G., Mwangi, W., Kimwele, M., & Gómez, J. M. (2023). A Cnn Model for Improved Image Denoising with an Attention Guided Feature Selection. 20 August 2023, PREPRINT (Version 1) https://doi.org/10.21203/rs.3.rs-3267082/v1

Mehta, D., Padalia, D., Vora, K., & Mehendale, N. (2022, December). *MRI image denoising using U-Net and Image Processing Techniques*. 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 306-313.

Mikołajczyk, A., & Grochowski, M. (2018, May). *Data augmentation for improving deep learning in image classification problem*. International interdisciplinary PhD workshop, Poland, 117-122.

Moghadam, F. S., & Rashidi, S. (2024). Novel feature extraction based on DCT-DOST features for classification of Digital Breast Tomosynthesis images into benign and malignant tumors. 07 February 2024, PREPRINT (Version 1), available at Research Square https://doi.org/10.21203/rs.3.rs-3931625/v1

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of big data*, **2**, 1-21 https://doi.org/10.1186/s40537-014-0007-7

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, et al. (2018). Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999

Paliwal, S. S., Vishwanath, D., Rahul, R., Sharma, M., & Vig, L. (2019, September). *Tablenet: Deep learning model for end-to-end table detection and tabular data extraction from scanned document images*. International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 128-133.

Patel, K. K., Goyal, S. K., & Patel, Y. K. (2023). Image Processing for Food Safety and Quality. Novel Technologies in Food Science, Navnidhi Chhikara, Anil Panghal, Gaurav Chaudhary (Editors), Wiley, 451-478.

Rafiee, A. A., & Farhang, M. (2023). A deep convolutional neural network for salt-and-pepper noise removal using selective convolutional blocks. *Applied Soft Computing*, **145**, 110535. https://doi.org/10.1016/j.asoc.2023.110535

Rashmi, R., Prasad, K., & Udupa, C. B. K. (2022). Breast histopathological image analysis using image processing techniques for diagnostic purposes: A

methodological review. *Journal of Medical Systems*, **46(7)**, https://doi.org/10.1007/s10916-021-01786-9

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: Convolutional networks for biomedical image segmentation*. 18th international conference on medical image computing and computer-assisted intervention, Munich, Germany, 234-241.

Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, **60(1-4)**, 259-268. https://doi.org/10.1016/0167-2789(92)90242-F

Salvi, M., Acharya, U. R., Molinari, F., & Meiburger, K. M. (2021). The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. *Computers in Biology and Medicine*, **128**, 104129. https://doi.org/10.1016/j.compbiomed.2020.104129

Sezer, A., & Altan, A. (2021). Detection of solder paste defects with an optimization-based deep learning model using image processing techniques. *Soldering & Surface Mount Technology*, **33(5)**, 291-298. https://doi.org/10.1108/SSMT-04-2021-0013

Srinivasa, K. G., Sowmya, B. J., Kumar, D. P., & Shetty, C. (2016). Efficient Image Denoising for Effective Digitization using Image Processing Techniques and Neural Networks. *International Journal of Applied Evolutionary Computation*, **7(4)**, 77-93. https://doi.org/10.4018/IJAEC.2016100105

Schreiber, S., Agne, S., Wolf, I., Dengel, A., & Ahmed, S. (2017). *Deepdesrt: Deep learning for detection and structure recognition of tables in document images*. 14th IAPR international conference on document analysis and recognition (ICDAR), Kyoto, Japan, 1162-1167.

Tahir, H., & Din, A. H. M. (2024). The Potential of Landsat 8 OLI Images in Coastline Identification: The Case Study of Basra, Iraq. *Engineering, Technology & Applied Science Research*, **14(1)**, 13041-13046. https://doi.org/10.48084/etasr.6580

Tomasi, C., & Manduchi, R. (1998). *Bilateral filtering for gray and color images*. Sixth international conference on computer vision, Bombay, India, 839-846.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008). *Extracting and composing robust features with denoising autoencoders*. In Proceedings of the 25th international conference on Machine learning, Helsinki, Finland, 1096-1103.

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. https://doi.org/10.1109/TIP.2003.819861

Yin, J., Xu, K., Kan, J., Dong, F., & Chen, K. (2023*). A Dual-U Structure for Image Denoising Based on Attention Mechanism.* 2nd International Conference on Computing, Communication, Perception and Quantum Technology (CCPQT), Xiamen, China, 421-426.

Zhao, R., Lun, D. P., & Lam, K. M. (2020). Enhancing and Learning Denoiser without Clean Reference. arXiv preprint arXiv:2009.04286. https://doi.org/10.48550/arXiv.2009.04286

Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, **26(7)**, 3142-3155. https://doi.org/10.1109/TIP.2017.2662206

Zhang, J., Niu, Y., Shangguan, Z., Gong, W., & Cheng, Y. (2023). A novel denoising method for CT images based on U-net and multi-attention. *Computers in Biology and Medicine*, **152**, 106387. https://doi.org/10.1016/j.compbiomed.2022.106387

Zhang, Q., Xiao, J., Tian, C., Chun-Wei Lin, J., & Zhang, S. (2023). A robust deformed convolutional neural network (CNN) for image denoising. *CAAI Transactions on Intelligence Technology*, **8(2)**, 331-342. https://doi.org/10.1049/cit2.12110

Zulkarnain, I., Nurmalasari, R. R., & Azizah, F. N. (2022). *Table information extraction using data augmentation on deep learning and image processing*. 16th International Conference on Telecommunication Systems, Services, and Applications, Lombok, Indonesia, 1-6.