https://doi.org/10.62836/jcmea.v3i1.030103

Global Science Publishing

JCMEA

Journal of Computational Methods in Engineering Applications

https://ojs.sgsci.org/journals/jcmea

Article

Deep Learning-Based Multifunctional End-to-End Model for Optical Character Classification and Denoising

Shuguang Xiong [1], Xiaoyang Chen [2] and Huitao Zhang [3,*]

[1] *Baidu Inc., Beijing, China; xiongshuguang@baidu.com*

[2] *Radiawave Co., Ltd., Shen Zhen, China; chenxiaoyang@radiawave.com*

[3] *Bybit Global Digital Solutions FZE, Dubai, United Arab Emirate*

Abstract: Optical Character Recognition (OCR) has revolutionized document processing by converting scanned documents, PDFs, and images captured by cameras into editable and searchable text. This technology is crucial for digitizing historical documents, streamlining data entry processes, and improving accessibility for the visually impaired through text-to-speech technologies. Despite its widespread application, OCR faces significant challenges, especially in accurately recognizing text in noisy or degraded images. Traditionally, OCR systems have treated noise reduction and character classification as separate stages, which can compromise the overall effectiveness of text recognition. Our research introduces a groundbreaking Multifunctional End-to-End Model for Optical Character Classification and Denoising, which integrates these functions within a unified framework. By employing a dual-output autoencoder, our model concurrently denoises images and recognizes characters, thereby enhancing both the efficiency and accuracy of OCR. This paper outlines the model's development and implementation, explores the interplay between denoising and classification, and presents compelling experimental results that demonstrate marked improvements over conventional OCR methods.

Keywords: component; Optical Character Classification; denoising; autoencoder; deep learning

1. Introduction

Optical Character Recognition (OCR) is a transformative technology in the digital text processing landscape, enabling the conversion of various types of documents—such as scanned paper documents, PDFs, or images captured by cameras—into editable and searchable data [1–3]. At its essence, OCR systems recognize and digitize characters in an image, converting them into a machine-readable text format. This technology has permeated numerous fields, including the digitization of historical documents, automation of data entry, and enhancement of accessibility for visually impaired individuals through text-to-speech systems. Moreover, advances in related fields, such as autonomous system optimization, directly contribute to enhancing the efficiency and accuracy of OCR technology [4]. The significance of OCR lies in its ability to dramatically reduce the manual labor involved in data entry while enhancing the accuracy and efficiency of information retrieval. For example, in libraries and archives, OCR facilitates the digital preservation of texts, making them easily accessible and searchable. Similarly, in biochemical analysis research, OCR efficiently and accurately processes and analyzes large volumes of data, thereby advancing the progress of precise biochemical

Received: 10 October 2023; Accepted: 31 October 2023.

* Corresponding: Huitao Zhang (lucker0087@gmail.com)

experiments [5, 6]. In the business realm, it streamlines processes such as invoice processing and receipt management, enabling quicker and more reliable extraction of information. Moreover, in the field of midinfrared photodetectors, such as those involving surface plasmon polariton graphene with multifrequency resonance, OCR aids in the management of vast amounts of data, ensuring that critical information is readily available when needed [7].

Despite its widespread applications and advancements, OCR technology still faces significant challenges. Chief among these is the accurate recognition of text in noisy or degraded images. Noise in images can arise from various sources, including poor lighting conditions, low-resolution scanning, or the presence of extraneous marks and stains on documents. Such noise can severely impact the performance of OCR systems, leading to errors in character recognition and ultimately compromising the reliability of the extracted text. To address these challenges, considerable research has focused on developing techniques for both text classification and image denoising. Text classification in OCR involves identifying and categorizing characters based on their visual features, while image denoising, enhanced by methods like semantic wireframe detection, aims to remove or reduce noise in images to improve the clarity and readability of the text [8]. Various methods, including deep learning models, have been employed to tackle these tasks independently due to their excellent performance in many tasks [9–15].

However, a notable gap exists in the current body of research: the integration of classification and denoising into a single, unified model. Most existing approaches treat these tasks separately, applying denoising techniques as a preprocessing step before feeding the cleaned images into classification models. This separation can lead to suboptimal performance, as the interdependencies between noise reduction and character recognition are not fully leveraged. For instance, an effective denoising process could remove features critical for accurate classification, while a classifier might misinterpret residual noise as part of the character structure.

Recognizing these limitations, our work proposes a novel approach: a Multifunctional End-to-End Model for Optical Character Classification and Denoising. This model aims to seamlessly integrate the denoising and classification processes, allowing for a more holistic and efficient approach to OCR. By building on the architecture of autoencoders, which are well-suited for denoising tasks, and enhancing them with classification capabilities, we seek to create a model that can simultaneously learn to clean noisy images and accurately recognize characters. The key innovation in our model lies in its dual-output structure, where the encoder extracts relevant features from the input images, and the decoder performs the denoising while a parallel classifier interprets the encoded features for character recognition. This design not only ensures that both tasks are optimized in tandem but also allows for the mutual reinforcement of denoising and classification objectives. The autoencoder's ability to reconstruct clean images provides a robust basis for the classifier, while the classifier's feedback helps refine the denoising process to preserve essential character details.

This paper is structured as follows: Section 2 details the related works on optical character classification and denoising shown in Figure 1. Following that, Section 3 provides the workflow of the proposed method in this study. The experimental results and corresponding discussion are presented in Section 4. Finally, Section 5 offers a comprehensive conclusion of this paper.

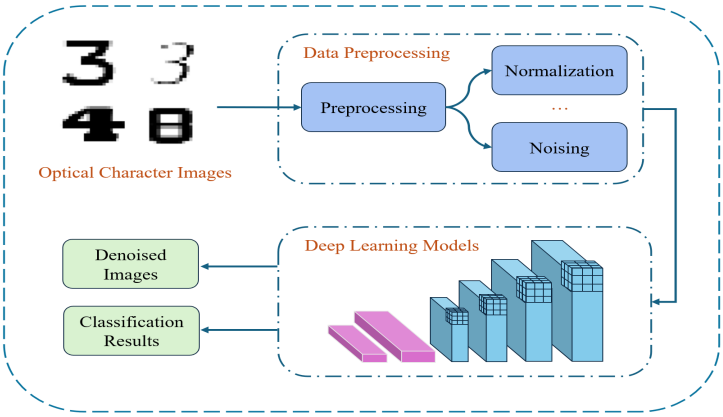

Figure 1. The general framework of deep learning-based Optical Character Classification and Denoising.

2. Literature Review

2.1. Optical Character Prediction

In terms of optical character detection and classification, many machine learning-based works have been carried out before since such performance have been demonstrated in many other studies [16–24]. For instance, Yi et al. developed a motion-based algorithm to identify text in cluttered environments, utilizing a method where the user shakes a hand-held object to detect interest points [25]. They employed Background Subtraction (BGR) to separate the moving object from the stationary background and trained the text classifier using the AdaBoost learning model [26]. Bai et al. introduced the Shared-Hidden Layer Deep Convolutional Neural Network (SHL-CNN), which reduces character recognition errors by 16-30% when trained for a single language and can be adapted for multilingual use [27]. Patil et al. and Shrivastava et al. detailed the phases in Optical Character Recognition, dividing the process into Image Processing and Recognition [28, 29]. P. Satyanarayana et al. researched assistive technologies for the blind, organizing the process into training, testing (classification using KNN), and text-to-speech conversion, with Raspberry Pi as the hardware interface [30]. Li et al. experimented with detecting text lines in handwritten documents by blurring the image horizontally multiple times, thresholding to connect words, and then grouping the thresholder image [31].

Shiravale et al. designed a segmentation technique that effectively separates the foreground from the background to enhance classification outcomes [32]. They employed a convolutional neural network encoder-decoder model to segment the background text, which was then recognized by the Tesseract OCR. Spyros et al. developed a meta-model known as the denoising autoencoder (DAE), which underwent Gaussian noise randomization as part of its graph neural network (GNN) architecture [33].

Several prior studies have also considered the denoising of optical character to further improve the performance. Ibrar et al. utilized nested denoising autoencoders, training them on multiple classes of ligatures from a noise-free Urdu printed text image (UPTI) dataset [34]. The trained network was then tested on a degraded version of the UPTI dataset, demonstrating accuracy between 93% and 96%. This performance surpasses that of traditional OCR methods for extensive ligature datasets. Alamsyah et al. focused on developing an autoencoder to denoise text images and evaluating the OCR performance in converting these denoised images into text [35]. Their research began with testing OCR on original and Gaussian noise-added text images. They then sought to optimize the autoencoder model by examining the effects of dataset size and optimizer type on denoising effectiveness. The culmination of their work involved evaluating OCR accuracy on text images denoised by the optimal autoencoder model. The study revealed that dataset size significantly impacts both denoising and OCR performance, with the autoencoder trained on a dataset size of 40 yielding the best results.

Although there has been significant progress in the fields of optical character classification and denoising, a notable gap remains in the current research: the integration of classification and denoising into a unified model. Most existing methods treat these tasks separately, using denoising as a preprocessing step before the cleaned images are processed by classification models. This separation may result in suboptimal performance due to the lack of full utilization of the interdependencies between noise reduction and character recognition. For example, a thorough denoising process might eliminate features essential for accurate classification, whereas a classifier could misconstrue residual noise as elements of the character structure.

3. Method

3.1. Dataset Preparation

In our study, the data utilized originates from Kaggle, a renowned platform for data science competitions, where users can find and publish datasets, explore and build models in a web-based data-science environment, and collaborate with other data enthusiasts. The dataset was generated using 3475 font styles available in Google Fonts, with each alphanumeric character (uppercase, lowercase, and numerals) produced in each font style and organized in a directory, totaling 2.1 lakh images across 62 classes. To balance performance and task complexity, we selected only 10 categories, specifically the digits 0-9. Each image is in grayscale, and the final dataset for training consists of 31, 257 images. These images were resized to 28x28 and normalized.

Normalization improves model training efficiency and helps in faster convergence by scaling the pixel values to a range, typically 0 to 1. The dataset was then split, with 80% used for training and 20% for testing. Some sample images collected are shown in Figure 2.

Figure 2. The sample images of the collected dataset.

To simulate the presence of noise in our images, we introduced Gaussian noise to both the training and testing datasets. This was done by adding noise with a mean of 0.0 and a standard deviation of 1.0, scaled by a noise factor of 0.1. By adjusting the noise factor, we controlled the intensity of the noise added to the images. This method allows us to create a more realistic scenario where the images are affected by typical disturbances, helping to train our model to be robust against such variations. Some sample noised images are provided in Figure 3.

Figure 3. The sample noised images of the collected dataset.

3.2. The Introduction of Deep Learning Models

3.2.1. CNN

Convolutional Neural Networks (CNNs) have revolutionized the field of computer vision and image processing [36–40], offering unparalleled accuracy in tasks such as image classification, object detection, and segmentation. This type of neural network is distinctively designed to automatically and adaptively learn spatial hierarchies of features, from low-level edges and textures to high-level patterns and object classes, through a series of convolutional layers.

The architecture of a CNN typically involves multiple layers, each playing a critical role in the extraction and transformation of features. The initial layers often consist of convolutional layers that apply various filters to the input image or feature maps from preceding layers. These filters are small but extend through the full depth of the input volume and are designed to capture basic features such as edges and simple textures. As the information progresses through the network, the layers become more adept at identifying complex patterns by combining the simpler features extracted by earlier layers. One of the significant advantages of CNNs is their ability to preserve the spatial relationship between pixels by learning image features using small squares of input data. This is unlike traditional neural networks that fully connect each neuron in one layer to every neuron in the next layer, leading to a drastic reduction in the number of parameters, which significantly enhances computational efficiency. This feature-sharing property allows CNNs to be deeper with fewer parameters compared to fully connected networks of the same size, enabling more complex and effective models to be trained.

Moreover, CNNs have been adapted for use in other domains beyond image processing, such as time series analysis, natural language processing, civil engineering and audio synthesis [41–43]. Additionally, OCR and federated learning models have advanced fields like cardiovascular medicine by identifying crucial amino acid residues in proteins and improving the estimation of tail risk measures in finance through approaches to extreme

value mixture modeling, respectively, demonstrating AI's broad impact across diverse scientific disciplines [44, 45]. In conclusion, the development of Convolutional Neural Networks (CNNs) marks a significant milestone in machine learning, combining architectural innovation with exceptional performance to tackle diverse digital challenges. Concurrently, research in distributed file systems has enhanced CNNs' potential in managing and optimizing complex data systems, reinforcing their capabilities in solving intricate problems.

3.2.2. Denoising Autoencoder

Denoising autoencoders are a transformative development in neural network technology [46], particularly adept at removing noise from data. These advanced autoencoders enhance model robustness and generalization by training them to recover pristine data from its noisy counterpart. By focusing on essential data features and disregarding extraneous noise, denoising autoencoders elevate performance across diverse tasks, including image restoration, signal processing, and anomaly detection.

Structured similarly to traditional autoencoders, denoising autoencoders consist of an encoder and a decoder. The encoder compresses the input into a condensed latent representation, capturing key data features, while the decoder reconstructs the data from this compressed state. However, denoising autoencoders introduce a twist: the input is intentionally contaminated with noise before being processed by the encoder. The model's training objective is to reconstruct the original, clean data from this corrupted input, compelling the autoencoder to distinguish between signal and noise.

The primary advantage of denoising autoencoders lies in their enhanced tolerance to input variations and perturbations. Trained with noisy data, these models excel at identifying and preserving the underlying data structure, even amidst significant noise. This attribute is invaluable in real-world scenarios where data is frequently imperfect, marred by photographic noise, sensor errors, or transmission disturbances.

3.3. Multifunctional End-to-End Model for Optical Character Classification and Denoising

In the proposed structure, the model integrates a denoising autoencoder with a classification component shown in Figure 4, allowing simultaneous output of denoised images and classification results. The architecture begins with the input image passing through convolutional layers with 32 filters of size 3x3 and ReLU activation, maintaining the spatial dimensions with same padding. This is followed by a max pooling layer that reduces the spatial dimensions while preserving essential features. The process is repeated with another set of convolutional layers, this time with 16 filters, and a subsequent max pooling layer to further compress the feature representation into a lower-dimensional encoded state. From this encoded state, the architecture branches off into two paths. The first path reconstructs the denoised image by using convolutional layers, starting with 16 filters and a ReLU activation, and then up-sampling to increase the spatial dimensions. This is followed by another convolutional layer with 32 filters and a final up-sampling step to match the original input dimensions. The reconstructed denoised image is produced by a convolutional layer with a single filter and a sigmoid activation, ensuring the output is in the same format as the input image. The second path focuses on classification. It begins by flattening the encoded representation to transform it into a vector. This vector is then processed by a dense layer with 128 units and ReLU activation to capture high-level features. The final output for classification is obtained through a dense layer with a number of units corresponding to the number of classes, using a softmax activation to provide probabilities for each class.

The training process optimizes both the denoising and classification tasks simultaneously by minimizing their combined loss functions. This integrated approach allows the model to leverage shared representations for both denoising and classification, potentially enhancing the performance of each task by learning from the synergies between them.

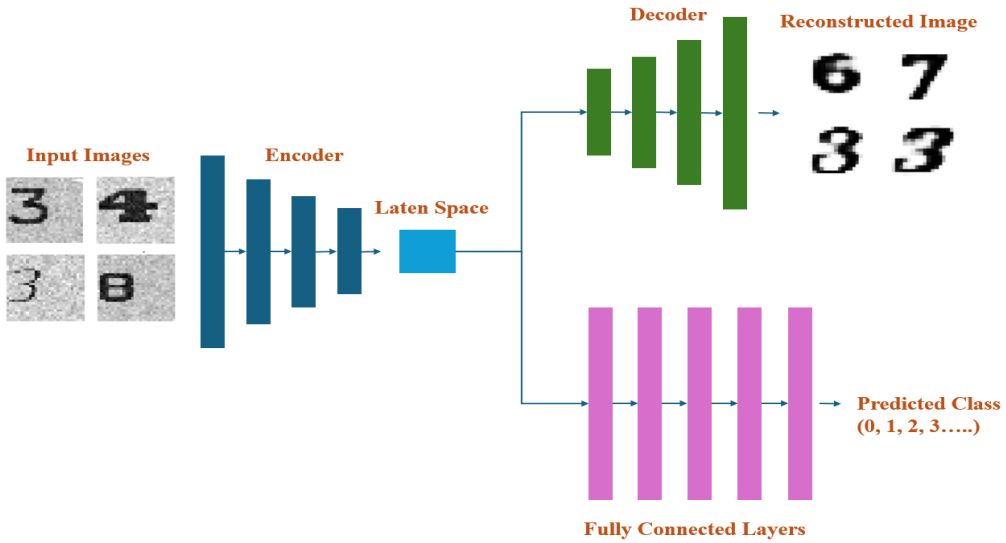

Figure 4. The architecture of the proposed autoencoder model combined with classification module.

3.4. Implementation Details

In this model configuration, the Adam optimizer is chosen for its efficiency in handling sparse gradients and adapting learning rates, which is particularly beneficial for convergence in complex models like this one. The loss function for the denoising part, measured by mean squared error (MSE), quantifies the difference between the original and the denoised images, encouraging the model to accurately reconstruct noise-free images. For the classification part, sparse categorical cross entropy is used, which allows the model to compute the loss directly without needing to convert the labels into a one-hot encoded format, simplifying the data preprocessing steps and enhancing computational efficiency. The losses from both the denoising and classification tasks are weighted equally, each contributing 50% to the total loss shown in equation (1). This balanced approach ensures that the model gives equal importance to both tasks during training, optimizing the shared features to perform well on both denoising and classification. The accuracy metric is applied to the classification output to monitor the model's performance in correctly predicting the class labels. During training, the model processes data in batches of 256 examples, which promotes efficient memory usage and expedites computation compared to handling individual samples. The training spans 50 epochs, providing ample iterations for the model to learn and refine its parameters, optimizing performance across both denoising and classification tasks. By shuffling the training data before each epoch, the model is prevented from learning any inherent order in the data, thereby enhancing its ability to generalize. Validation occurs with the noisy test set and its corresponding clean images and labels, offering a continuous evaluation of the model's performance on unseen data throughout the training process.

$$L_{total} = 0.5\, L_{MSE} + 0.5\, L_{Crossentropy} \tag{1}$$

4. Results and Discussion

4.1. The Performance of the Classification

The training performance of the model is effectively captured through three key graphs, each demonstrating significant insights into the model's learning dynamics. The first graph shown in Figure 5, illustrating the total loss, shows a sharp initial decline followed by a more gradual decrease over successive epochs. This steep initial drop indicates the model's rapid adaptation to the major patterns within the dataset, thereby significantly improving its predictive capabilities from the outset. As training progresses, the reduction in loss slows, suggesting that the model is fine-tuning its parameters to capture more subtle features in the data.

The second graph shown in Figure 6 displays the classification loss, which follows a similar trajectory to the total loss—characterized by a sharp decline early in training and leveling off to a steadier, slower decrease. This pattern underscores the efficiency of the classification component of the model, which quickly learns to

distinguish between different classes and incrementally refines its predictions as training advances.

The third graph shown in Figure 7, depicting the classification accuracy, shows a rapid increase to high levels early in the training process, eventually stabilizing above 90%. This swift improvement in accuracy demonstrates the model's capability to effectively learn and adapt to the classification task, achieving and maintaining high accuracy consistently throughout the training.

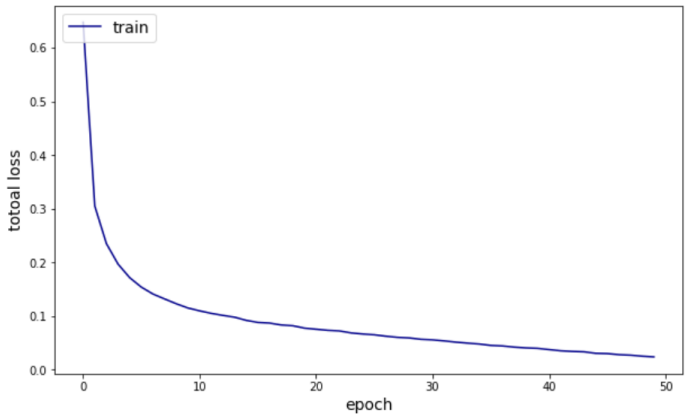

Figure 5. The total loss curve of the proposed model during the training.

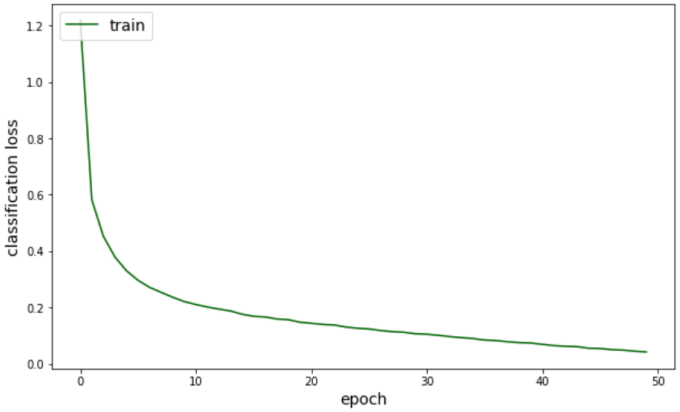

Figure 6. The classification loss curve of the proposed model during the training.

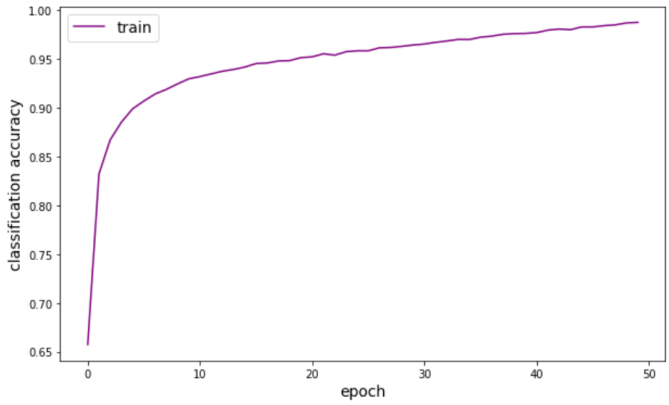

Figure 7. The classification accuracy curve of the proposed model during the training.

The results presented in Table 1 and Figure 8 demonstrate the high classification performance of the proposed model on the testing dataset. According to Table 1, the model achieves an accuracy of 93.17%, with similarly high precision and recall scores, both at 93.87%. The F1 score, which is a harmonic mean of precision and recall, also stands impressively at 93.80%, indicating a well-balanced model that efficiently manages the trade-off between precision and recall.

Figure 8 provides a detailed view of the model's performance across individual classes through the confusion matrix. This matrix shows the number of predictions for each class compared to the true labels. For most classes, the diagonal elements, which represent the number of correct predictions, are significantly higher than the off-diagonal elements, suggesting that the model has a high rate of correct classifications for each class. For instance, the model correctly predicted the class '0' 625 times, whereas misclassifications for this class are relatively low, indicating strong recognition capability. The confusion matrix also reveals some areas of confusion between specific classes, although these instances are relatively few. For example, class '9' was occasionally confused with class '7', as indicated by 16 misclassifications. However, these numbers are small compared to the correct predictions, underscoring the model's overall effectiveness.

Table 1. The classification performance of the proposed model in testing dataset.

Metrics	Value
Accuracy	0.9317
Precision	0.9387
Recall	0.9383
F1 score	0.9380

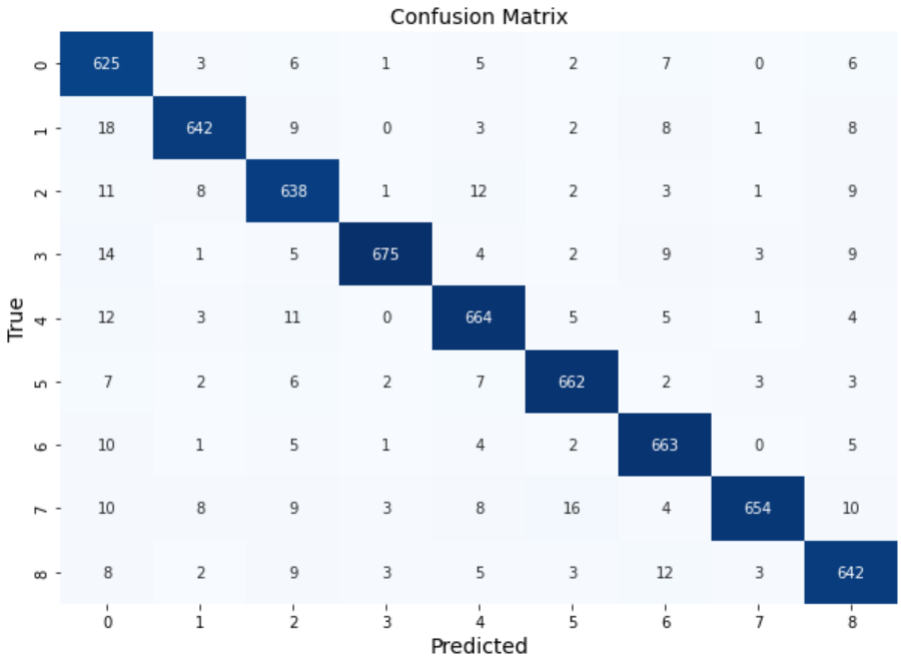

Figure 8. The confusion matrix based on the proposed model.

4.2. The Performance of the Denoising

Figure 9 presents the reconstruction loss curve of the proposed model during the training. Initially, the graph demonstrates a sharp decrease in loss, dropping from around 0.07 to below 0.02 within the first few epochs. This rapid decline indicates that the denoising model quickly adapted to the major features required for noise reduction in the data, achieving significant improvements in reconstructing the clean data from noisy inputs early in the training process. After this initial steep decrease, the reconstruction loss levels off and stabilizes, maintaining a value just above 0.01 for the remainder of the epochs. The flattening of the curve suggests that the model reached a point of diminishing returns where further improvements in learning the noise characteristics and reconstructing the clean data became incremental. The low and stable reconstruction loss after the initial drop reflects the model's effectiveness in denoising, with the fine-tuning phases solidifying the gains in noise reduction capabilities.

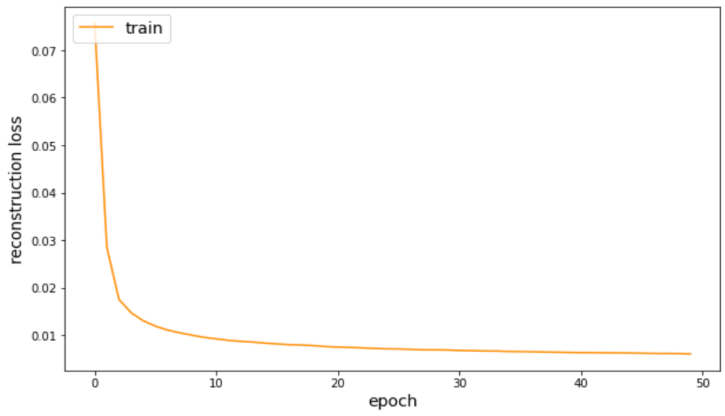

Figure 9. The reconstruction loss curve of the proposed model during the training.

Some sample denoised images are provided in Figure 10. The denoising performance of the model is outstanding, effectively restoring images to a quality that closely matches the originals. This exceptional ability to reduce noise while preserving the critical details and characteristics of the original images highlights the robust denoising capabilities of the autoencoder. These results not only confirm the effectiveness of the autoencoder in eliminating noise but also demonstrate its practical utility in scenarios where maintaining the integrity of the original data is essential.

Figure 10. The sample denoised images provided by the autoencoder.

4.3. Discussion

The experimental results clearly demonstrate that integrating denoising and classification tasks within a single autoencoder model is not only feasible but also highly effective. This dual-task approach allows the model to leverage shared representations, which enhances learning efficiency and performance across both tasks. The training curves, as depicted in Figures 5 through 7, reveal a rapid initial decrease in both total and classification losses, indicating that the model swiftly adapts to the key features of the dataset. This quick learning phase is critical as it shows the model's ability to grasp essential patterns that contribute to both denoising and classification effectively. Despite these promising results, there are inherent limitations to consider. The performance of the model is highly dependent on the quality and variety of the training data. In scenarios where the data diversity is limited or if the noise patterns are highly irregular, the model might not generalize well to unseen data types or noise levels. Additionally, the computational demand of simultaneously optimizing for denoising and classification can be significant, potentially limiting the model's scalability or efficiency when deployed in resource-constrained environments. Future research could explore more dynamic training strategies that adaptively weigh the loss contributions from denoising and classification tasks to further enhance model performance. Investigating the impact of different types of noise and more complex classification tasks might also provide deeper insights into the model's capabilities and limitations. Expanding the model's robustness to handle a broader range of noise patterns and data variations will be crucial for its adoption in more diverse real-world applications. In addition, the hardware conditions can be also further improved using some new technologies [47–53].

5. Conclusion

The integration of denoising and classification tasks within a unified autoencoder model has been notably

successful. Our model not only showcases its viability but also delivers significant enhancements in OCR performance, processing noisy images efficiently while ensuring high accuracy in character recognition. The experimental results affirm the model's robust performance, highlighted by exceptional accuracy, precision, and recall, culminating in a balanced F1 score of 93.80%. However, the effectiveness of the model is contingent on the quality and diversity of the training data, and it places considerable demands on computational resources. Future efforts should aim to improve the model's ability to adapt to various noise patterns and develop dynamic training approaches that more effectively fine-tune both denoising and classification tasks. Enhancing the model's robustness and scalability is essential for its application in more diverse real-world environments, positioning it as a more adaptable tool in the realm of digital processing.

Funding

Not applicable.

Author Contributions

Conceptualization, writing—original draft preparation and writing—review and editing, S.X., X.C. and H.Z.; All of the authors read and agreed to the published the final manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Chaudhuri A, *et al*. *Optical Character Recognition Systems*; Springer: Berlin, Germany, 2017.

2 Eikvil L. Optical Character Recognition. Research Report, NorskRegnesentral, Blindern, 1993; **26**.

3 Nagy G, Nartker TA, Rice SV. Optical Character Recognition: An Illustrated Guide to the Frontier. *Document Recognition and Retrieval VII* 1999; **3967**: 58–69.

4 Sun G, Zhan T, Owusu BG, Daniel A–M, Liu G, Jiang W. Revised Reinforcement Learning Based on Anchor Graph Hashing for Autonomous Cell Activation in Cloud–RANs. *Future Generation Computer Systems* 2020; **104**: 60–73.

5 Horne J, Beddingfield E, Knapp M, Mitchell S, Crawford L, Mills SB, Wrist A, Zhang S, Summers RM. Caffeine and Theophylline Inhibit β–Galactosidase Activity and Reduce Expression in Escherichia coli. *ACS Omega* 2020; **5**(**50**): 32250–32255.

6 Mock MB, Zhang S, Pniak B, Belt N, Witherspoon M, Summers RM. Substrate Promiscuity of the NdmCDE N7–Demethylase Enzyme Complex. *Biotechnology Notes* 2021; **2**: 18–25.

7 Deng X, Kawano Y. Surface Plasmon Polariton Graphene Midinfrared Photodetector with Multifrequency Resonance. *Journal of Nanophotonics* 2018; **12**(**2**): 026017–026017.

8 Zhou Y, Osman A, Willms M, Kunz A, Philipp S, Blatt J, Eul S. Semantic Wireframe Detection. In Proceedings of the DACH-Jahrestagung 2023, Friedrichshafen, Germany, 15–17 May 2023.

9 Liu Y, Liu L, Yang L, Hao L, Bao Y. Measuring Distance Using Ultra–Wideband Radio Technology Enhanced by Extreme Gradient Boosting Decision Tree (XGBoost). *Automation in Construction* 2021; **126**:

103678.

10 Yu F, Milord J, Orton S, Flores L, Marra R. Students' Evaluation Toward Online Teaching Strategies for Engineering Courses during COVID. In Proceedings of the 2021 ASEE Midwest Section Conference, Virtual, 13–15 September 2021.

11 Liu Y, Bao Y. Real–Time Remote Measurement of Distance Using Ultra–Wideband (UWB) Sensors. *Automation in Construction* 2023; **150**: 104849.

12 Liu Y, Bao Y. Review of Electromagnetic Waves–Based Distance Measurement Technologies for Remote Monitoring of Civil Engineering Structures. *Measurement* 2021; **176**: 109193.

13 Chen H, Chen P, Qiu Y, Ling N. FARNet: Fragmented Affinity Reasoning Network of Text Instances for Arbitrary Shape Text Detection. *IET Image Processing* 2023; **17**(**6**): 1959–1977.

14 Wu P, Liu A, Fu J, Ye X, Zhao Y. Autonomous Surface Crack Identification of Concrete Structures Based on an Improved One–Stage Object Detection Algorithm. *Engineering Structures* 2022; **272**: 114962.

15 Wang B, Shen Y, Zhai L, Xia X, Gu H-M, Wang M, Zhao Y, Chang X, Alabi A, Xing S, Deng S, Liu B, Wang G, Qin S, Zhang D-W. Atherosclerosis–Associated Hepatic Secretion of VLDL but Not PCSK9 Is Dependent on Cargo Receptor Protein Surf4. *Journal of Lipid Research* 2021; **62**: 100091.

16 Zhao F, Yu F, Trull T, Shang Y. A New Method Using LLMs for Keypoints Generation in Qualitative Data Analysis. In Proceedings of the 2023 IEEE Conference on Artificial Intelligence (CAI), Santa Clara, CA, USA, 5–6 June 2023.

17 Hao Y, Chen Z, Jin J, Sun X. Joint Operation Planning of Drivers and Trucks for Semi–Autonomous Truck Platooning. *Transportmetrica A: Transport Science* 2023; 1–37. DOI: 10.1080/23249935.2023.2266041.

18 Frid–Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN–Based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification. *Neurocomputing* 2018; **321**: 321–331.

19 Yu F, Milord J, Orton SL, Flores L, Marra R. The Concerns and Perceived Challenges Students Faced When Traditional in–Person Engineering Courses Suddenly Transitioned to Remote Learning. In Proceedings of the 2022 ASEE Annual Conference, Minneapolis, MN, USA, 26–29 June 2022.

20 Yu F, Strobel J. Work–in–Progress: Pre–college Teachers' Metaphorical Beliefs about Engineering. In Proceedings of the 2021 IEEE Global Engineering Education Conference (EDUCON), Vienna, Austria, 21–23 April 2021.

21 Milord J, Yu F, Orton S, Flores L, Marra R. Impact of COVID Transition to Remote Learning on Engineering Self–Efficacy and Outcome Expectations. In Proceedings of the 2021 ASEE Virtual Annual Conference, Virtual Conference, 26–29 July 2021.

22 Li S, Singh K, Riedel N, Yu F, Jahnke I. Digital Learning Experience Design and Research of a Self–Paced Online Course for Risk–Based Inspection of Food Imports. *Food Control* 2022; **135**: 108698.

23 Yu F, Milord JO, Flores LY, Marra R. Work in Progress: Faculty Choice and Reflection on Teaching Strategies to Improve Engineering Self–Efficacy. In Proceedings of the 2022 ASEE Annual Conference, Minneapolis, MN, USA, 26–29 June 2022.

24 Orton S, Yu F, Flores L, Marra R. Student Perceptions of Confidence in Learning and Teaching Before and After Teaching Improvements. In Proceedings of the 2023 ASEE Annual Conference, Baltimore, MD, USA, 25–28 June 2023.

25 Yi C, Tian Y, Arditi A. Portable Camera–Based Assistive Text and Product Label Reading from Hand–Held Objects for Blind Persons. *IEEE/ASME Transactions On Mechatronics* 2013; **19**(**3**): 808–817.

26 Freund Y, Schapire RE. Experiments with a New Boosting Algorithm. In Proceedings of the Proceedings of the Thirteenth International Conference (ICML '96), Bari, Italy, 3–6 July 1996.

27 Bai J, Chen Z, Feng B, Xu B. Image Character Recognition Using Deep Convolutional Neural Network Learned from Different Languages. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014.

28 Patil V, Sanap RV, Kharate RB. Optical Character Recognition Using Artificial Neural Network. *Int J Eng Res Gen Sci* 2015; **3**(**1**): 7.

29 Shrivastava V, Sharma N. Artificial Neural Network Based Optical Character Recognition. 2012. arXiv: 1211.4385.

30 Satyanarayana P, Sujitha K, Sai Anitha Kiron V, Ajitha Reddy P, Ganesh M. Assistance Vision for Blind People Using K–NN Algorithm and Raspberry Pi. In *Proceedings of the 2nd International Conference on Micro–Electronics, Electromagnetics and Telecommunications: ICMEET 2016*; Springer: Berlin, Germany, 2018.

31 Li Y, Zheng Y, Doermann D. Detecting Text Lines in Handwritten Documents. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006.

32 Shiravale SS, Jayadevan R, Sannakki SS. Recognition of Devanagari Scene Text Using Autoencoder CNN. *ELCVIA: Electronic Letters on Computer Vision and Image Analysis* 2021; **20**(**1**): 0055–69.

33 Gidaris S, Komodakis N. Generating Classification Weights with Gnn Denoising Autoencoders for Few–Shot Learning. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.

34 Ahmad I, Wang X, Li R, Rasheed S. Offline Urdu Nastaleeq Optical Character Recognition Based on Stacked Denoising Autoencoder. *China Communications* 2017; **14**(**1**): 146–157.

35 Alamsyah N, Fauzan MN, Putrada AG, Pane SF. Autoencoder Image Denoising to Increase Optical Character Recognition Performance in Text Conversion. In Proceedings of the 2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS), Bandung, Indonesia, 23 November 2022.

36 Qiu Y, Wang J, Jin Z, Chen H, Zhang M, Guo L. Pose–Guided Matching Based on Deep Learning for Assessing Quality of Action on Rehabilitation Training. *Biomedical Signal Processing and Control* 2022; **72**: 103323.

37 Deng X, Simanullang M, Kawano Y. Ge–Core/a–Si–Shell Nanowire–Based Field–Effect Transistor for Sensitive Terahertz Detection. *Photonics* 2018; **5**(**2**): 13.

38 Liu Y, Yang H, Wu C. Unveiling Patterns: A Study on Semi–Supervised Classification of Strip Surface Defects. *IEEE Access* 2023; **11**: 119933–119946.

39 Liu Y, Bao Y. Automatic Interpretation of Strain Distributions Measured from Distributed Fiber Optic Sensors for Crack Monitoring. *Measurement* 2023; **211**: 112629.

40 Ismail WN, Hassan MM, Alsalamah HA, Fortino G. CNN–Based Health Model for Regular Health Factors Analysis in Internet–of–Medical Things Environment. *IEEE Access* 2020; **8**: 52541–52549.

41 Kayalibay B, Jensen G, van der Smagt P. CNN–Based Segmentation of Medical Imaging Data. 2017. arXiv: 1701.03056.

42 Salehi AW, *et al*. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. *Sustainability* 2023; **15**(**7**): 5930.

43 Wenjun D, Fatahizadeh M, Touchaei HG, Moayedi H, Foong LK. Application of Six Neural Network–Based Solutions on Bearing Capacity of Shallow Footing on Double–Layer Soils. *Steel and Composite Structures* 2023; **49**(**2**): 231–244.

44 Wang M, *et al*. Identification of Amino Acid Residues in the MT–Loop of MT1–MMP Critical for Its Ability to Cleave Low–Density Lipoprotein Receptor. *Frontiers in Cardiovascular Medicine* 2022; **9**: 917238.

45 Qiu Y. *Estimation of Tail Risk Measures in Finance: Approaches to Extreme Value Mixture Modeling*; Johns Hopkins University: Baltimore, MD, USA, 2019.

46 Qiu Y, Yang Y, Lin Z, Chen P, Luo Y, Huang W. Improved Denoising Autoencoder for Maritime Image Denoising and Semantic Segmentation of USV. *China Communications* 2020; **17**(**3**): 46–57.

47 Deng X, Oda S, Kawano Y. Graphene–Based Midinfrared Photodetector with Bull'S Eye Plasmonic Antenna. *Optical Engineering* 2023; **62**(**9**): 097102–097102.

48 Sugaya T, Deng X. Resonant Frequency Tuning of Terahertz Plasmonic Structures Based on Solid Immersion Method. In Proceedings of the 2019 44th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW–THz), Paris, France, 1–6 September 2019.

49 Tao G, et al. Surf4 (Surfeit Locus Protein 4) Deficiency Reduces Intestinal Lipid Absorption and Secretion and Decreases Metabolism in Mice. *Arterioscler Thromb Vasc Biol* 2023; **43**(**4**): 562–580.

50 Shen Y, Gu H – m, Zhai L, Wang B, Qin S, Zhang D – w. The Role of Hepatic Surf4 in Lipoprotein Metabolism and the Development of Atherosclerosis in apoE−/− mice. *Biochimica et Biophysica Acta (BBA) –Molecular and Cell Biology of Lipids* 2022; **1867**(**10**): 159196.

51 Deng X, Li L, Enomoto M, Kawano Y. Continuously Frequency – Tuneable Plasmonic Structures for Terahertz Bio–Sensing and Spectroscopy. *Scientific Reports* 2019; **9**(**1**): 3498.

52 Gu Y, Sharma S, Chen K. Demo: Image Disguising for Scalable GPU – accelerated Confidential Deep Learning. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, in CCS '23. New York, NY, USA: Association for Computing Machinery, Copenhagen, Denmark, 26–30 November *2023*.

53 Gu Y, Chen K. GAN – Based Domain Inference Attack. *Proceedings of the AAAI Conference on Artificial Intelligence* 2023; **37**(**12**): 14214–14222.