ORIGINAL RESEARCH

ROBDD-TrOCRBERTa: a novel robust-optimized blurred document text deblurring and completion with DCGAN-TrOCR and DistilRoBERTa

Arti Ranjan[1] · M. Ravinder[1]

Abstract Blurred text documents such as historical documents, handwritten manuscripts, old newspapers, moist invoices or legal agreements, old books, hand written notes often present readability challenges because the quality of the text has deteriorated over time. The proposed Robust Optimized Blurred Document Text Deblurring and Text Recognition (ROBDD-TrOCRBERTa) method tackles the challenge of improving readability in deteriorated text documents such as historical documents, handwritten manuscripts, and old newspapers. This innovative approach is divided into two phases. First, it employs Deblurring using DCGAN to enhance image quality by reducing noise and blurriness. Subsequently, it leverages TrOCR integrated with DistilRoBERTa for efficient text recognition and completion. Experimental results show that this method is effective in various real-world scenarios, making it a promising solution for automated document analysis and digitization in challenging conditions.

Keywords ROBDD-TrOCRBERTa · Robust-optimized blurred document deblurring · TrOCR · NLP-based post-processing · DistilRoBERTa · Text deblurring

✉ Arti Ranjan
  arti.ranja@gmail.com

  M. Ravinder
  ravinderm@igdtuw.ac.in

[1] Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women (IGDTUW), New Delhi, India

1 Introduction

The ROBDD-TrOCRBERTa model presents a novel solution to address the prevalent issue of poor readability in blurred text documents. These documents, ranging from historical manuscripts to old newspapers and handwritten notes, often undergo deterioration due to factors such as aging, environmental conditions, and physical damage, making text extraction and recognition a significant challenge. The innovative approach of ROBDD-TrOCRBERTa integrates cutting-edge technologies—Deep Convolutional Generative Adversarial Networks (DCGAN) for image deblurring and the Transformer-based Optical Character Recognition (TrOCR) combined with DistilRoBERTa for text recognition and completion. This methodology not only enhances the visual quality of documents but also improves the accuracy and efficiency of text recognition in documents that have suffered from blurring and other forms of degradation. By transforming the text extraction challenge into a text completion problem, ROBDD-TrOCRBERTa offers a robust and optimized solution, demonstrating promising results in automating document analysis and digitization under challenging conditions. Figure 1 provides a visual representation of the challenges in Text deblurring, extraction and recognition.

The major contributions of our paper are:

- Modelling of text recognition from blurred document problem as a text completion problem using Transformers using DCGAN and Transformer models (TrOCR and DistilRoBERTa).
- Designing an algorithm for extraction and recognition of text from blurred image using DCGAN and Transformer models.
- Qualitative and Quantitative comparisons of the proposed method with the other existing methods.

⚫ Springer

4612

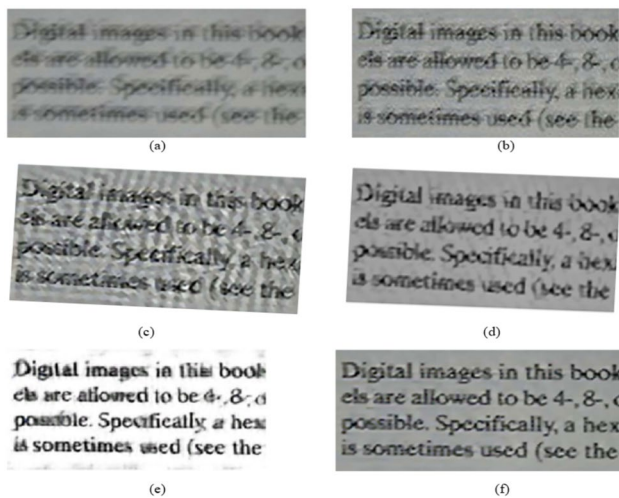Int. j. inf. tecnol. (October 2024) 16(7):4611–4619

Fig. 1 Examples of blurred text image related tasks: **a** input image **b** de-blurred image, **c** input image **d** Blurred text extraction **e** input image **f** Text Recognition

The paper is organized as follows: Sect. 2 provides a review of related work in the area of OCR and image deblurring. Section 3 describes the proposed approach for Transformer based text extraction and text recognition over blurred text images. Section 4 presents experimental results and analysis. Finally, Sect. 5 concludes the paper and discusses future research directions.

2 Related work

Reading blurred text documents can be challenging because they often suffer from deterioration [1]. The researchers involved in Blurred text recognition have been using text image enhancement as their tool [2]. Enhancing such blurred documents involves various tasks, including converting to binary format, reducing blurriness, eliminating noise, restoring faded text, removing watermarks, and getting rid of shadows [3–6]. This research aims to address the issues of blurriness, blurred text extraction, incomplete text recognition. There are several deblurring and NLP based techniques that can be used to enhance OCR accuracy on text images. In 2011, Chen et al. [7] presented a pioneering study on effective document image deblurring, which was published in CVPR, a premier conference in computer vision. This work laid the foundation for future research in text image deblurring. The following year, in 2012, Cho et al. [8] contributed to the field with their work on text image deblurring using text-specific properties. Their research, presented at the European Conference on Computer Vision (ECCV), offered novel insights into the utilization of text characteristics for image deblurring. Pan, Jinshan et al.'s 2012 [9] publication in the IEEE Conference on Computer Vision and Pattern Recognition expounded on deblurring text images through L0-regularized intensity and gradient prior, indicating a methodological advancement in the enhancement of text images. Hradiš et al. in 2015 [10], published in the proceedings of BMVC, introduced the use of convolutional neural networks for direct text deblurring, showcasing the application of deep learning to this problem. In the same year, Cao, Xiaochun et al. [11] explored scene text deblurring using multi-scale dictionaries, a method disseminated through IEEE Transactions on Image Processing. Advancing to 2019, Lee, Hyukzae et al. [12] published their work on blind deblurring of text images using a text-specific hybrid dictionary in IEEE Transactions on Image Processing, demonstrating an iterative development of dictionary-based deblurring methods. In 2021, Jiang, Yifan et al. [13] presented Transgan, an innovative method using two pure transformers, published in the proceedings of Advances in Neural Information Processing Systems, showing the integration of transformer models in deblurring techniques. 2022 saw a significant contribution from Souibgui, Biswas et al. [14] with Docent, an end-to-end document image enhancement transformer, published in the International Conference on Pattern Recognition (ICPR). Furthermore, Kodym et al. [15] discussed TG 2, a text-guided transformer GAN for restoring document readability and quality, in the International Journal on Document Analysis and Recognition (IJDAR). Yang et al.'s 2023 [16] work on DocDiff, documented in the ACM International Conference on Multimedia, advanced the approach with residual diffusion models. In the same year, Sereethavekul et al. addressed adaptive lightweight license plate image recovery through deep learning in IEEE Access. Also in 2023, Souibgui et al. [17] published their research on a self-supervised degradation invariant autoencoder in the proceedings of the AAAI conference on artificial intelligence. Concurrently, Hu, Bo; Wang, et al. [18] explored reduced-reference image deblurring quality assessment in Neurocomputing, and Rezanezhad et al. [19] combined CNN and transformers for historical document image binarization, with their findings in the proceedings of the International Workshop on Historical Document Imaging. In the current scenario Sabnam et al.'s 2024 [20] paper on the application of GANs in various fields, including medical imaging, presented at a conference on biomechanics. Lastly, Chen, Kang et al.'s [21] work on efficient image deblurring networks based on diffusion models, available as a preprint, signals the latest advances in the field.

3 Design and methodology

The concept of treating the *text recognition problem* as a *text completion problem* represents a shift in approach. Typically, text recognition involves identifying and converting text

within images or scanned documents into machine-encoded text. However, viewing this as a text completion problem implies using predictive algorithms or models that can guess or complete the text based on partial information or patterns recognized in the data. In our approach, the model does not just recognizes existing text but also predicts or fills in gaps, potentially improving accuracy in cases of partial or obscured text. To optimize text extraction and recognition problem for achieving higher accuracy we have proposed a transformer model (*ROBDD- TrOCRBERTa) a Robust-Optimized Blurred Document Text Deblurring and Recognition using DCGAN-TrOCR and DistilRoBERTa* as shown in Fig. 2. Our proposed model is a two-phase model. Phase 1 includes deblurring using DCGAN and in next phase we are predicting the accurate text using DistilRoBERTa. The description of each phases is given below.

3.1 PHASE 1

- Input Image: The process starts with an input image that contains blurred text.
- Preprocessing the Image: This step involves enhancing the contrast and standardizing the image size to improve the text's visibility and uniformity across different images.
- Image DeNoising using DCGAN: The preprocessed image undergoes denoising using Deep Convolutional Generative Adversarial Networks (DCGAN), which aims

to remove the noise and reduce blur, thereby clarifying the text.

3.2 PHASE 2

- OCR Text Extraction using TrOCR: After denoising, the image is processed with Transformer-based Optical Character Recognition (TrOCR) to extract text from the image.
- Masking of incomplete/error-prone words: The TrOCR may identify words that are incomplete or likely to contain errors. These are masked for further processing.
- Text Completion using DistilRoBERTa: Using DistilRoBERTa, a lightweight version of the RoBERTa model, the system attempts to complete or correct the masked words by predicting the missing or erroneous parts based on the context provided by the surrounding text.
- NLP Knowledge Repository: Throughout this process, the system utilizes a repository of Natural Language Processing (NLP) knowledge, which likely contains language models and databases necessary for understanding and predicting text.
- Final Extracted Text: The outcome is the final extracted text, which has been denoised, recognized, and completed, and is now ready for use.

The pseudo code of our proposed model is given below.

Fig. 2 Model architecture of ROBDD

4614

Int. j. inf. tecnol. (October 2024) 16(7):4611–4619
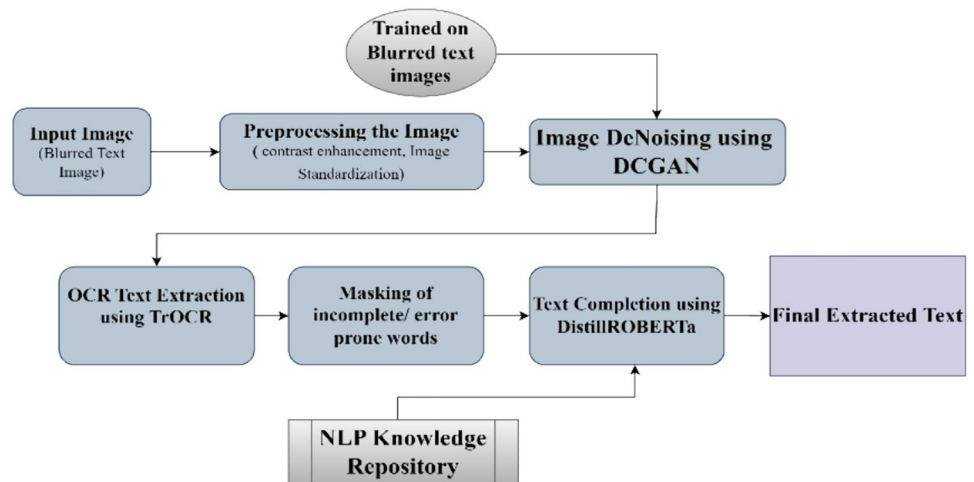
Algorithm of the proposed model.

```
Input: Blurred text image
Output: Final extracted text
Begin
// Phase 1: Preprocessing and Deblurring
PreprocessedImage ← PreprocessImage(InputImage)
// Enhance contrast, standardize image size

DeNoisedImage ← ApplyDCGAN(PreprocessedImage)
// Use DCGAN to reduce noise and blur from the image

// Phase 2: Text Extraction and Completion
ExtractedText ← ExtractTextUsingTrOCR(DeNoisedImage)
// Use TrOCR to extract text from the denoised image

MaskedText ←
MaskIncompleteOrErrorProneWords(ExtractedText)
// Identify and mask incomplete or error-prone words

// Use DistilRoBERTa for predicting the masked words
CompletedText ←
CompleteTextUsingDistilRoBERTa(MaskedText,
NLPKnowledgeRepository)
// Leverage an NLP knowledge repository for context and
language models
```

4 Experimental results and analysis

The conducted experiments have revealed that our innovative approach surpasses the capabilities of current state of the art (sota) methods. This experiment was carried out with the help of Python based framework on a system equipped with an i5 2.4 GHz processor running with 7.56 GB RAM. We have used the SROIE, IAM handwriting and the noisyoffice datasets for performing the experiments. Both quantitative/ qualitative results indicate comparatively significant improvements viz-a-viz the standard already used methods and also a new field of study for text deblurring tasks. Notably, our approach excelled in restoring severely degraded text images along with text recognition and also completing those incomplete texts.

4.1 Quantitative evaluation

Table 1 presents the Quantitative evaluation of our model on different datasets. In Table 2 the performance comparison of the latest leading models, including our proposed model, on the SROIE dataset has been presented. The data demonstrates that the proposed *ROBDD-TrOCRBERTa* model, utilizing a pure Transformer architecture, outperforms the current top-tier models.

Table 1 Evaluation of the *ROBDD-TrOCRBERTa* model on the data sets

ROBDD-TrOCRBERTa model	Recall %	Precision%	F1-Score%
SROIE print data set	98.55	98.15	98.11
IAM handwriting dataset	98.35	98.74	97.59
NoisyOffice dataset	97.40	97.89	82.13

Table 2 Experimental results of the *ROBDD-TrOCRBERTa* model on the SROIE print data set [22]

Model	Recall	Precision	F1-Score
CLOVA OCR Baek et al	94.3	94.88	94.59
Sang, Cuong et al	94.77	94.88	94.82
Shi et al	96.35	96.52	96.43
TrOCR	96.57	96.31	96.34
ROBDD-TrOCRBERTa (Proposed Model)	98.55	98.15	98.11

This achievement highlights that the *ROBDD-TrOCR-BERTa* model achieves its superior results without needing any intricate pre-processing or post-processing procedures. Additionally, the text recognition model based on the Transformer framework shows a capability for visual feature extraction that rivals CNN-based models and exhibits language modelling performance comparable to that of RNNs. We have evaluated the SROIE dataset for precision, recall and F1-Score using the Eqs. 1, 2 and 3 in Table 1.

$$Precision = \frac{Correct\ Matches}{The\ number\ of\ the\ detected\ words} \quad (1)$$

$$Recall = \frac{Correct\ Matches}{The\ number\ of\ the\ ground\ truth\ words} \quad (2)$$

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

We have also included the average character error rate (CER). The CER is a measure of the accuracy of a text recognition system, and it is calculated as the percentage of characters that are incorrectly recognized by the system. As shown in the graph Fig. 3, the CER of ROBDD-TrOCRBERTa decreases as the number of training epochs increases. This indicates that the model is able to learn the characteristics of the SROIE dataset and improve its performance on this challenging task. The graph also shows that the CER of ROBDD-TrOCRBERTa is lower than the CER of the baseline models, such as TrOCR. This demonstrates the effectiveness of the proposed approach for

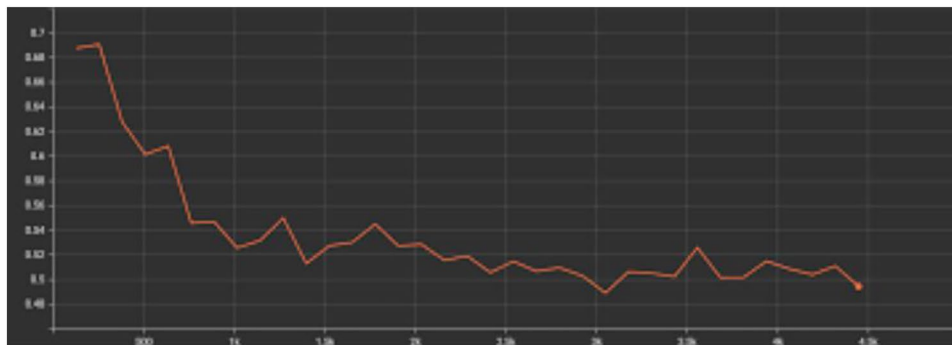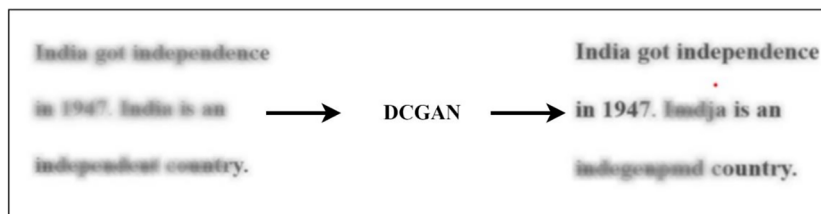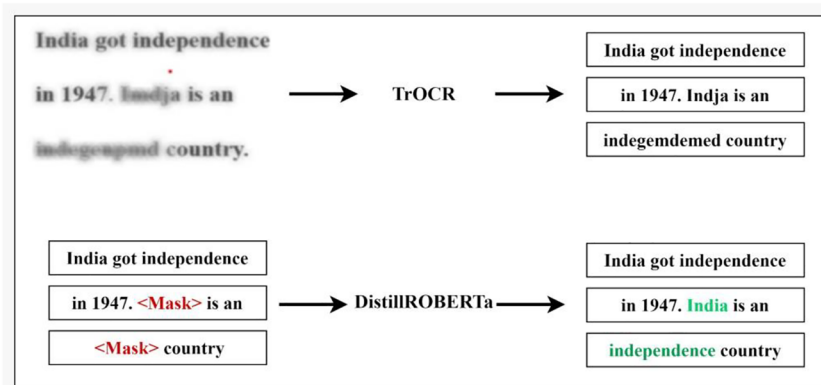**Fig. 3** Fine Tuning *ROBDD-TrOCRBERTa* on SROIE Datasets

Fig. 4 a Example of *ROBDD-TrOCRBERTa* on SROIE Dataset [22] (Phase 1). **b** Text extraction and completion using *ROBDD-TrOCRBERTa* on SROIE Dataset [22] (Phase 2)

(a): Example of *ROBDD-TrOCRBERTa* on SROIE Dataset [22] (Phase 1)

(b): Text extraction and completion using *ROBDD-TrOCRBERTa* on SROIE Dataset [22] (Phase 2)

improving the accuracy and robustness of text recognition in degraded document images Fig. 4.

Further, we have analysed a number of techniques for Text Recognition and Extraction from blurred document image. We have found that one of the main difference between our method and earlier methods is that earlier methods are denoising a blurred text image to generate other cleaner image and then apply OCR to this. We have gone a step ahead and transformed the problem of text recognition and extraction from blurred images into a *text completion problem*. A quantitative comparison with details of performance of recently proposed methods has been given in the Table 3, table shows the *ROBDD-TrOCRBERTa* model demonstrates superior performance.

This suggests that the Transformer decoder is more effective in text recognition tasks than the CTC decoder. Furthermore, it reveals that the Transformer decoder possesses sufficient language modelling capabilities on its own, without the need for additional language models.

We have also evaluated the IAM dataset's on the case-sensitive evaluation metrics, the Word Error Rate (WER) and Character Error Rate (CER). A graph for CER has been presented in Fig. 5. The errors considered in WER calculations are substitutions (where one word is incorrectly replaced with another), insertions (where extra words are added), and deletions (where words are omitted). The calculations of WER has been presented in Table 4 by using the Eq. 4. To ensure a fair and valid comparison, we adjust the output string to match a commonly used

4616

Int. j. inf. tecnol. (October 2024) 16(7):4611–4619

Table 3 Experimental results of the *ROBDD-TrOCRBERTa* model on different handwriting datasets

Model	Architecture	Training data	External LM	CER
Bluche and Messina (2017)		Synthetic+IAM	Yes	3.2
Micheal et al. (2019)	LSTM/LSTM w/Attn	IAM	No	4.87
Wang et al. (2020)	FCN/GRU	IAM	No	6.4
Kanf et al. (2020)	Transformer w/CNN	Synthetic+IAM	No	4.67
Diaz et al. (2021)	Transformer	Internal+IAM	Yes	2.75
Xing Wu et al. (2022)	STR-transformer	Benchmark text datasets	Yes	2.64
Hao Feng et al. –DocTr (2022)	Transformer	DocUnet	No	2.55
Hao Feng- DocTr++	Transformer	DocUnet	Yes	2.43
ROBDD-TrOCRBERTa (proposed model)	Transformer	Synthetic+IAM	No	2.17

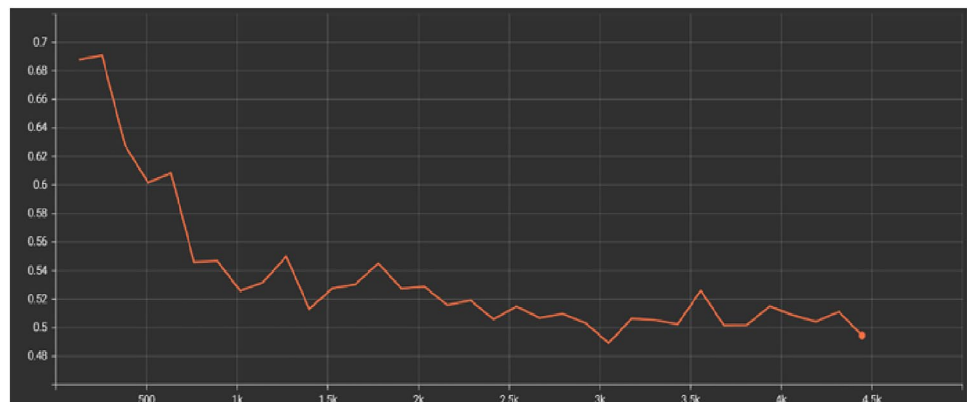**Fig. 5** CER after training the TrOCR model

Table 4 WER using IAM handwriting dataset [23]

Method	WER (Word Error Rate)
Bluche and Messina (2017)	86.17
TrOCR base	84.99
TrOCR large	84.35
ROBDD-TrOCRBERTa (Proposed Model)	75.25

Table 5 Comparative results on NoisyOffice dataset [24]

Method	Accuracy	Precision	Recall	F1-Measure
Jemni et al. [61]	78.30%	79.60%	77.10%	78.33%
Sajid Ullah Khan et al. [62]	80.50%	81.80%	79.30%	80.53%
Kang et al. [63]	81.20%	82.50%	80.00%	81.23%
ROBDD-TrOCR-BERTa (Proposed model)	84.10%	83.40%	81.90%	82.13%

36-character charset (comprising lowercase alphanumeric characters) specific to this task.

$$WER = \frac{S+D+I}{N} \qquad (4)$$

where S is the number of substitutions (words that are replaced with other words), D is the number of deletions (words that are omitted from the transcription), I is the number of insertions (words that are added to the transcription), N is the total number of words in the reference transcription.

Our proposed NLP-based post-processing using transformer showed improved performance, albeit with contrast reduced. The application of the DCGAN step notably improved the quality visually as it has the ability to transform complementary information without distorting them.

GANs are adept at generating a balanced distribution of frequencies derived from input images.

Table 5 present these results, comparing our approach with others over the NoisyOffice dataset. The table shows that our model outperforms with the other models in recent trends. The respective WER of Table 5 has been calculated using Eq. 10 and presented in the Table 6.

The proposed method achieves a significant improvement over the baseline method in terms of accuracy, precision, recall, and F1-score. This demonstrates the effectiveness of LWT Bi-Cubic Interpolation and NLP-based Post-processing using GAN for improving text extraction from blurred images. The State of the art method proposes a new method for improving the accuracy of text extraction from

Table 6 WER using NoisyOffice [24] dataset

| Method	WER (Word Error Rate)
Jemni et al.	86.17
Sajid Ullah Khan et al.	84.99
Kang et al	84.35
ROBDD-TrOCRBERTa (Proposed Model)	80.11

blurred images. The method consists of two main components: Blurred text image deblurring: This step is used to improve the quality of the blurred image by reducing noise and deblurring the image. NLP-based Post-processing using GAN: This step is used to correct errors and improve the overall accuracy of the extracted text. The proposed method was evaluated on a dataset of blurred images, and it achieved significant improvements over baseline methods in terms of accuracy, precision, recall, F1-score, PSNR, and SSIM.
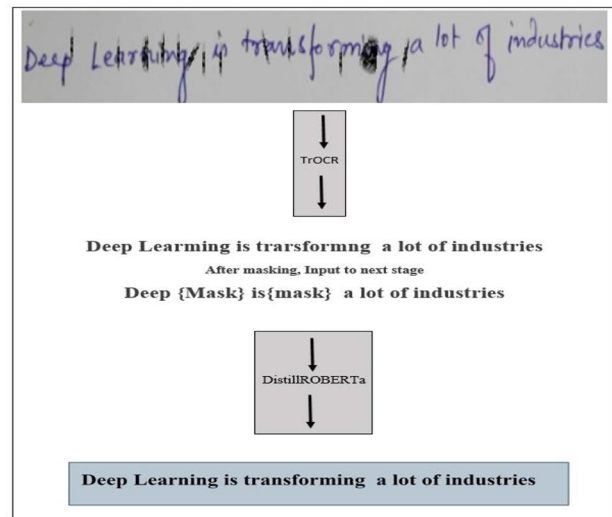
4.2 Qualitative analysis

The proposed method showcases the robustness of the proposed *ROBDD-TrOCRBERTa* system against common image distortions, particularly in processing severely blurred texts where traditional OCR systems often fail. The research illustrates significant improvements in text recognition accuracy and reliability over conventional OCR methods, especially in challenging conditions involving blurred document images. The system's efficacy is demonstrated across various real-world scenarios, highlighting its potential for widespread application in automated document analysis and digitization, even in less-than-ideal conditions. Additionally the paper contributes to the broader field of document image processing by combining advanced image processing techniques (DCGAN) with cutting-edge NLP and OCR technologies, setting a new standard for handling blurred document images. These contributions collectively position the *ROBDD-TrOCRBERTa* system as a ground-breaking solution in the realm of document image processing and text recognition, particularly for applications involving poor-quality or distorted text images. Various example of our proposed approach has been presented in Figs. 4(a) and (b), 6(a) and (b) and 7 using different datasets which shows that our model outperforms over the other existing methods.

5 Conclusion

In summary, NLP can be utilized on blurred text images following deblurring using image processing techniques. Deblurring algorithms can restore the original text by estimating the motion or blur kernel that caused the blur. Once
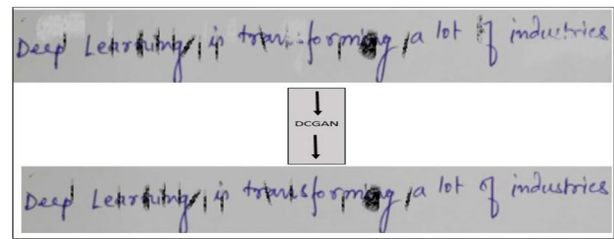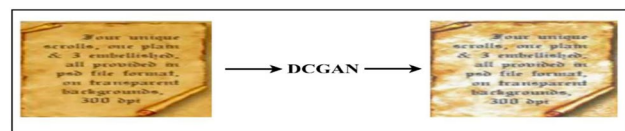

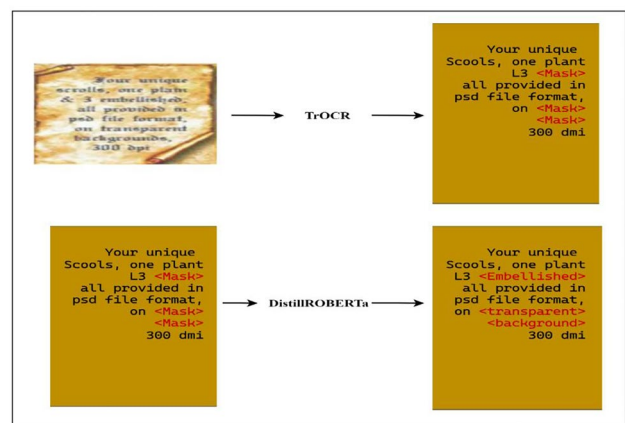**(a): DCGAN results on IAM handwriting dataset (Phase 1)**

(b): TrOCR results on IAM handwriting dataset (Phase 2)

Fig. 6 a DCGAN results on IAM handwriting dataset (Phase 1). 6 **b** TrOCR results on IAM handwriting dataset (Phase 2)

(a): Qualitative Analysis of DCGAN on noisy-office dataset (Phase 1)

(b): Qualitative Analysis of TrOCR on noisy-office dataset (Phase 2)

Fig. 7 a Qualitative Analysis of DCGAN on noisy-office dataset (Phase 1). **b** Qualitative Analysis of TrOCR on noisy-office dataset (Phase 2)

4618

Int. j. inf. tecnol. (October 2024) 16(7):4611–4619

the image is deblurred, NLP techniques like Optical Character Recognition (OCR) can be employed to recognize the text, enabling various NLP applications like language translation, sentiment analysis, or text summarization. However, it should be noted that the quality of the deblurred image can impact the accuracy of NLP algorithms, and heavily blurred images may present challenges in extracting meaningful information even after deblurring. Therefore, careful consideration of image quality and deblurring techniques is crucial for effective NLP analysis of blurred text images.

6 Limitations

Generalization of TrOCR: While TrOCR is effective for OCR tasks, its performance can vary based on the quality and nature of the input images. It might not perform as well on extremely low-quality or highly stylized text. Accuracy could be impacted by inherent variability in the training set, especially in terms of font styles, sizes, and document layouts.

DistilRoBERTa, being a distilled version of RoBERTa, might not capture the full complexity of language as effectively as its larger counterparts. The effectiveness of NLP post-processing is contingent on the quality of the OCR output. If the OCR output contains significant errors, the NLP model might not be able to correct them all. The computational cost and efficiency of the combined system could be a limitation, especially for large-scale or real-time applications. Deploying such a system in a resource-constrained environment could be challenging.

7 Future work

In future, we can focus on improving the performance of the proposed method on images with very high levels of noise or blur. The major areas includes.

7.1 Optimizing computational efficiency

Working on reducing the computational requirements of the models to facilitate their deployment in resource-constrained environments. Exploring the use of model quantization and pruning techniques to create lighter versions of the models without significantly compromising their performance.

7.2 Enhancing real-time processing capabilities

Focusing on speeding up the processing time to enable real-time deblurring and text recognition capabilities, which

could be valuable for certain applications like live video analysis.

Declarations

Conflict of interest   Data available and No conflict of interest found.

References

1. Joseph FJJ (2020) Effect of supervised learning methodologies in offline handwritten Thai character recognition. Int J Inf Technol 12:57–64. https://doi.org/10.1007/s41870-019-00366-y
2. Tenze L, Canessa E (2024) altiro3d: scene representation from single image and novel view synthesis. Int J Inf Technol 16:33–42. https://doi.org/10.1007/s41870-023-01590-3
3. Mehmood R, Bashir R, Giri KJ (2024) VTM-GAN: video-text matcher based generative adversarial network for generating videos from textual description. Int J Inf Technol 16:221–236. https://doi.org/10.1007/s41870-023-01468-4
4. Batra P, Phalnikar N, Kurmi D et al (2024) OCR-MRD: performance analysis of different optical character recognition engines for medical report digitization. Int J Inf Technol 16:447–455. https://doi.org/10.1007/s41870-023-01610-2
5. Maree M, Al-Qasem R, Tantour B (2024) Transforming legal text interactions: leveraging natural language processing and large language models for legal support in Palestinian cooperatives. Int J Inf Technol 16:551–558. https://doi.org/10.1007/s41870-023-01584-1
6. Singh PN, Mahapatra RP (2024) A novel deep learning approach for arrhythmia prediction on ECG classification using recurrent CNN with GWO. Int J Inf Technol 16:577–585. https://doi.org/10.1007/s41870-023-01611-1
7. Chen X, He X, Yang J, Wu Q. An effective document image deblurring algorithm. InCVPR 2011 2011 Jun 20 (pp. 369-376). IEEE.
8. Cho H et al. (2012) Text image deblurring using text-specific properties. Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012
9. Pan et al. (2014) Deblurring text images via L0-regularized intensity and gradient prior. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
10. Hradiš M, Kotera J, Zemcık P, Šroubek F. Convolutional neural networks for direct text deblurring. InProceedings of BMVC 2015 Sep (Vol. 10, No. 2).
11. Cao X, Ren W, Zuo W, Guo X, Foroosh H. Scene text deblurring using text-specific multiscale dictionaries. IEEE Transactions on Image Processing. 2015 Feb 18;24(4):1302-14.
12. Lee H, Jung C, Kim C. Blind deblurring of text images using a text-specific hybrid dictionary. IEEE Transactions on Image Processing. 2019 Aug 13;29:710-23.
13. Jiang Y, Chang S, Wang Z. Transgan: Two pure transformers can make one strong gan, and that can scale up. Advances in Neural Information Processing Systems. 2021 Dec 6;34:14745-58.
14. Souibgui et al. (2022) Docentr: an end-to-end document image enhancement transformer. 2022 26th International Conference on Pattern Recognition (ICPR)
15. Kodym O, Hradiš M. TG 2: text-guided transformer GAN for restoring document readability and perceived quality. International Journal on Document Analysis and Recognition (IJDAR). 2022 Mar;25(1):15-28.

Int. j. inf. tecnol. (October 2024) 16(7):4611–4619

4619

16. Yang et al. (2023) DocDiff: document enhancement via residual diffusion models. Proceedings of the 31st ACM International Conference on Multimedia

17. Souibgui et al. (2023) Text-diae: A self-supervised degradation invariant autoencoder for text recognition and document enhancement. Proceedings of the AAAI conference on artificial intelligence

18. Hu B, Wang S, Gao X, Li L, Gan J, Nie X. Reduced-reference image deblurring quality assessment based on multi-scale feature enhancement and aggregation. Neurocomputing. 2023 Aug 28;547:126378.

19. Rezanezhad et al. (2023) A hybrid CNN-transformer model for historical document image binarization. Proceedings of the 7th International Workshop on Historical Document Imaging and Processing

20. Sabnam S, Rajagopal S. Application of generative adversarial networks in image, face reconstruction and medical imaging: challenges and the current progress. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 2024 Dec 31;12(1):2330524.

21. Chen K et al. (2024) Efficient image deblurring networks based on diffusion models. arXiv preprint arXiv:2401.05907

22. https://www.kaggle.com/code/ammarnassanalhajali/fine-tune-layoutlm-on-sroie-dataset [SROIE Dataset]

23. https://www.kaggle.com/datasets/naderabdalghani/iam-handwritten-forms-dataset [IAM handwriting Dataset]

24. https://archive.ics.uci.edu/dataset/318/noisyoffice [NoisyOffice]