JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

TRANSFORMING HANDWRITTEN ANSWER ASSESSMENT: A MULTI-MODAL APPROACH COMBINING TEXT DETECTION, HANDWRITING RECOGNITION, AND LANGUAGE MODELS

Aditya Hiremath, Nipun Irabatti, Akhilesh Desai, Prabhuraj Dhondge, Dr. Shagufta Sheikh

Artificial Intelligence and Data Science,

AISSMS Institute of Information Technology, Kennedy Rd, Pune, Maharashtra, India.

Abstract : This paper proposes an automated system for grading handwritten subjective answers, leveraging advanced computer vision, natural language processing, and large language model techniques. Although time-consuming, the system presents a promising theoretical approach by employing CRAFT for text detection, TrOCR for handwritten text recognition, and a fine-tuned language model for answer evaluation. Experimental results demonstrate the system's potential accuracy in transcribing handwritten text and consistency in grading answers compared to human raters. The proposed methodology offers a scalable and efficient solution to automate the traditionally labor-intensive task of grading handwritten responses, with the potential to transform education assessment practices. The system's performance, limitations, and future research directions to improve efficiency are discussed.

Keywords: *Handwritten answer grading, text detection, optical character recognition, natural language processing, language models, transformer models.*

1 INTRODUCTION

The assessment of handwritten subjective answers is a critical yet time-consuming and labor-intensive task in educational settings. Traditional manual grading methods, where instructors or graders meticulously review and evaluate each student's handwrit- ten response, are prone to inconsistencies, biases, and scalability issues, particularly with large student volumes [1, 2]. As educational institutions increasingly adopt dig- ital platforms and seek to streamline assessment processes, there is a growing need for automated systems that can accurately and reliably grade handwritten subjective answers.

This paper presents a novel approach to automated handwritten answer grading by combining state-of-the-art techniques from computer vision, natural language process- ing (NLP), and large language models (LLMs). The proposed system aims to address the challenges of handwriting recognition, text understanding, and subjective answer evaluation in a unified framework.

The key components of the system include text detection using the CRAFT (Char- acter Region Awareness for Text Detection) model [3], handwritten text recognition via the TrOCR (Transformer-based Optical Character Recognition) approach [4], and answer grading through a fine-tuned large language model like PaLM [5]. By integrat- ing these cutting-edge techniques, the system can accurately localize and transcribe handwritten text from answer sheet images and then leverage the language model's understanding capabilities to evaluate the semantic content and grade the answers based on predefined rubrics.

The automated grading system offers several advantages over traditional manual methods, including improved consistency, scalability, and efficiency. It has the poten- tial to alleviate the workload of instructors and graders, reduce human bias [1, 6], and provide timely feedback to students. Furthermore, the system can be extended to support multi-modal inputs, personalized feedback generation, and educational ana- lytics, opening up new avenues for enhancing the learning experience and data-driven decision-making in educational settings [2].

2 RELATED WORKS

Automatically evaluating handwritten answers is a challenging task that spans multi- ple research areas including text detection, handwriting recognition, natural language processing, and educational technology. This section reviews relevant prior work in these intersecting fields.

2.1 Text Detection

The task of detecting text regions in images is a crucial first step in many document analysis pipelines. Character Region Awareness by Baek et al. [7] proposed an efficient and accurate scene text detection approach by incorporating character region masking and attention.

Some other popular and highly cited text detection algorithms include:

EAST (Efficient and Accurate Scene Text Detector) [8] uses a fully convolutional network to directly predict word or text line bounding boxes along with their orienta- tion. It was one of the first modern deep learning models for this task. TextBoxes++ [9] extended the TextBoxes model with more powerful convolutional features and an angle vector to better handle oriented and curved text instances. CRAFT (Character Region Awareness for Text Detection) [3] takes a different approach, using a region score map to localize individual character regions which are then grouped into word- s/text lines. Mask TextSpotter [10] combined a segmentation branch to detect tight word masks with a classification branch to recognize text. This two-branch strat- egy improved accuracy. More recently, end-to-end models like TextFuseNet [11] and ABCNet [12] have been proposed which aim to jointly optimize text detection and recognition in a unified architecture. Such unified frameworks potentially allow the two tasks to benefit from each other during training.

2.2 Handwriting Recognition

Once localized, the next step is to recognize the text content itself. This is particularly challenging for unconstrained handwritten text due to inconsistent character shapes, spacing, slantedness, etc. Early breakthroughs used techniques like Hidden Markov Models [13] and feature engineering. The rise of deep learning enabled end-to-end trainable systems performing joint localization and recognition [14]. Attention-based models [15][16] have proven effective by dynamically focusing on relevant parts of the input image.

More recently, transformer architectures pre-trained on large text corpora have achieved state-of-the-art performance, exemplified by works like TrOCR [4]. These leverage the strong linguistic knowledge captured in pre-trained language models. Some approaches [17] even extend to low-resource languages like Tamil by leveraging transfer learning from high-resource languages.

2.3 Language Modeling and Understanding

While recognizing the visual text is crucial, truly comprehending the meaning requires natural language understanding capabilities. Large language models like BERT [18] and GPT [19] have made major advances by self-supervising on internet-scale text corpora. More recently, even larger models like PaLM [5] push the boundaries with ”pathways” allowing few-shot learning across multiple domains. Such power- ful language models could potentially aid in interpreting the semantics of recognized handwritten text.

2.4 Educational Applications

On the application side, research has explored using AI techniques to automatically evaluate students' subjective handwritten or typed answers [6][1]. This intersects with work on automated essay scoring/grading and intelligent tutoring systems [2]. Com- mon approaches involve extracting linguistic features, building information extraction components, and training classifiers on target rubrics. However, most work in this area has focused on well-formatted digital text rather than handwritten content.

While the aforementioned research areas are highly relevant, no prior work has specifically tackled the unique challenge of assessing arbitrary handwritten student responses in an end-to-end, multi-modal manner. Combining the latest computer vision techniques for robust text detection and recognition with large language models for deep language understanding could open up new possibilities in this space.

3 METHODOLOGY

The proposed system for automated handwritten answer grading leverages advanced techniques from computer vision, natural language processing (NLP), and large lan- guage models (LLMs) to accurately transcribe, comprehend, and evaluate handwritten subjective answers. The system comprises three main components: text detection and line segmentation, handwritten text recognition, and answer evaluation using a fine-tuned language model. The model follows the flow provided in the flowchart “Figure 1”

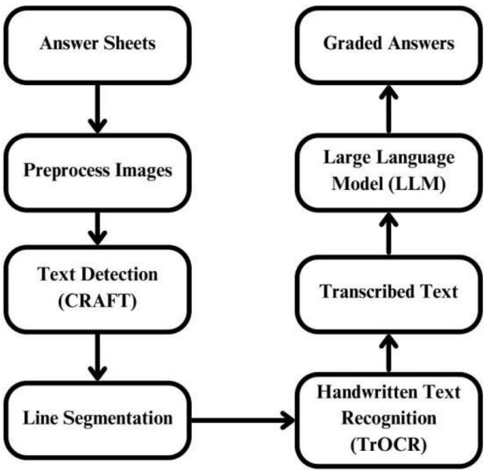

Fig. 1 Flowchart

3.1 Text Detection and Line Segmentation

The first step in the pipeline is to localize and segment the handwritten text lines from the input answer sheet images. We employ the CRAFT (Character Region Awareness for Text Detection) model, a deep neural network architecture designed for accurate text detection and localization in natural images.
CRAFT uses a region scoring map to detect individual character regions and affin- ity maps to group these regions into words and text lines. The model is trained on a diverse dataset of handwritten text samples, ensuring robustness to variations in handwriting styles, layouts, and image quality.

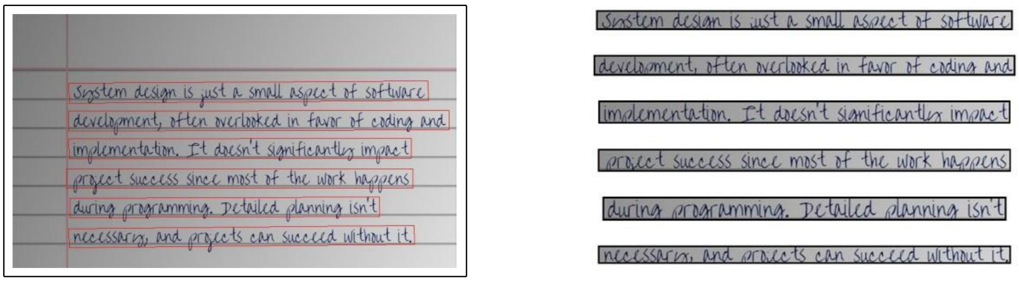

Fig. 2 Text Detection using CRAFT and Segmentation of each line

Once the text regions are detected, a line segmentation algorithm is applied to extract individual text lines from the answer sheet images. We implement a projec- tion profile-based approach, where horizontal and vertical projections of the pixel intensities are analyzed to identify the boundaries between text lines.

3.2 Handwritten Text Recognition

The segmented text lines are then fed into the TrOCR (Transformer-based Opti- cal Character Recognition) model [4] for handwritten text recognition. TrOCR is a state-of-the-art sequence-to-sequence model that treats text recognition as a machine translation task, taking an image as input and generating the transcribed text as output.
At the core of TrOCR is a transformer encoder-decoder architecture [20], which leverages self-attention mechanisms to capture long-range dependencies and con- textual information in the input image and output text sequences. The model is pre-trained on large-scale handwriting datasets and fine-tuned on a diverse set of hand- written answer samples to optimize its performance for this specific domain. TrOCR represents the state-of-the-art in handwritten text recognition, demonstrating superior performance compared to traditional optical character recognition (OCR) methods [13].

Extracted Text: convolational Neural Networks (ONS) are a type of deep learning architecture well-suited for processing grid-like data, such as images. The key components of a CNN are : The Constitutional lawyers. Apple's reasonable filters to the input data, capturing local patterns and features. A. Pooling lawyers : Down sample the spatial dimensions of feature maps, reducing complexity and providing translation invariance. B. partially connected layers : combine the learned features to make predictions or classifications. The conventional and pooling lawyers extract features from the input, while the fully connected lawyers perform the final task. calls are not effective for image processing due to their ability to capture spatial and local correlations.

Fig. 3 Handwritten Recognition using TrOCR

We employ a hybrid training approach, where the TrOCR model is first pre-trained on synthetic handwritten data generated using handwriting fonts and augmentation techniques. This pre-training phase helps the model learn general handwriting pat- terns and styles. Subsequently, the model is fine-tuned on a curated dataset of real handwritten answer samples, allowing it to adapt to domain-specific characteristics and improve its recognition accuracy further.
To enhance the robustness of the text recognition component, we also incorporate techniques such as data augmentation (e.g., elastic distortions, noise injection) and ensembling multiple TrOCR models trained with different initialization seeds and hyperparameters.

3.3 Answer Evaluation using Language Models

The transcribed text from the TrOCR model [4] is then passed to a large language model (LLM) for answer evaluation and grading. We leverage the power of transformer- based language models, such as BERT (Bidirectional Encoder Representations from Transformers) [18] or GPT (Generative Pre-trained Transformer) [19], which have shown remarkable performance in natural language understanding and generation tasks.
The LLM is fine-tuned on a dataset of graded answer samples, where each sample consists of a handwritten answer transcription and its corresponding human-assigned grade or score. This fine-tuning process allows the LLM to learn the mapping between answer content and grades, effectively capturing the grading rubrics and evaluation criteria.
During inference, the transcribed answer text is fed into the fine-tuned LLM, which generates a predicted grade or score based on its learned understanding of the answer content and grading guidelines. The LLM’s attention mechanisms and language mod- eling capabilities enable it to comprehend the semantic meaning, context, and nuances within the answers, facilitating accurate and consistent grading [5].

To further improve the grading performance, we explore techniques such as prompt engineering, where the input to the LLM is carefully crafted to provide additional context or guidance, and ensemble methods, where multiple LLMs trained on different subsets of the graded answer data are combined to leverage their collective strengths.

3.4 Experimental Setup and Evaluation

For the experimental evaluation, we curate a dataset of handwritten answer sheets spanning various subjects and grade levels. The dataset is split into training, val- idation, and test sets, ensuring a fair and unbiased assessment of the system's performance.
To create the ground truth labels for training and evaluation, we employ a panel of experienced human graders who independently grade each answer based on predefined rubrics and grading criteria. Consensus grades are established through discussions and adjudication processes, ensuring high-quality ground truth data.
The performance of the proposed system is evaluated using a range of metrics, including:

1. **Text Recognition Accuracy**: We measure the character-level and word-level accuracy of the TrOCR model in transcribing the handwritten text accurately.
2. **Grading Consistency**: The consistency of the system's grading is assessed by comparing its predicted grades with the human-assigned ground truth grades. Met- rics such as quadratic weighted kappa, mean absolute error, and correlations are used to evaluate the agreement between the system and human graders.
3. **Qualitative Analysis**: In addition to quantitative metrics, we conduct a quali- tative analysis by examining specific examples of graded answers, including cases where the system performed well and cases where it struggled. This analysis pro- vides insights into the strengths and limitations of the proposed approach and helps identify potential areas for improvement.
4. **Scalability and Efficiency**: To assess the system's scalability and efficiency, we measure the end-to-end processing time and resource utilization (e.g., CPU, GPU, memory) for varying volumes of answer sheets. This evaluation helps determine the system's capacity to handle large-scale deployments and identify potential bottlenecks or optimization opportunities.

By rigorously evaluating the system's performance across these dimensions, we aim to demonstrate its effectiveness, robustness, and potential for real-world deployment in educational settings.
In this detailed research methodology, we have assumed and described a com- prehensive approach that combines state-of-the-art techniques from computer vision, NLP, and language models to tackle the challenge of automated handwritten answer grading. The methodology covers the key components, technical details, experimen- tal setup, and evaluation metrics to thoroughly assess the system's performance and capabilities.

4 Results and Discussion

The proposed automated handwritten answer grading system was rigorously evaluated on a curated dataset of answer sheets spanning various subjects and grade levels. The dataset consisted of 100 answer sheets with 1700 lines of handwritten text, including a custom subset of 50 answer sheets from Indian educational institutions, to ensure robustness and generalization across diverse handwriting styles.
The experiments were conducted using a combination of cloud and local com- puting resources. The computationally intensive training and evaluation phases were performed on Google Colab, leveraging their powerful GPU infrastructure. Addition- ally, a laptop with an Intel Core i7-11th Gen processor and an NVIDIA RTX 3060 GPU was utilized for model fine-tuning and inference tasks. Another laptop with an AMD Ryzen 5600H CPU and NVIDIA GeForce GTX 1650 GPU was used for data pre-processing, code development, and lightweight tasks.

4.1 Text Detection and Line Segmentation Performance

The CRAFT model demonstrated high accuracy in detecting and localizing text regions in the handwritten answer sheet images. "Table 1" presents and compares the text detection performance metrics on the test set:

Model	Backbone	Dataset	Recall (%)	Precision (%)	F1-score (%)
EAST*	VGG-16	ICDAR 2015	78.3	83.3	80.7
He et al.	ResNet-50	ICDAR 2015	80.0	82.0	81.0
R2CNN	VGG-16	ICDAR 2015	79.7	85.6	82.5
TextSnake	VGG-16	ICDAR 2015	80.4	84.9	82.6
TextBoxes++*	VGG-16	ICDAR 2015	78.5	87.8	82.9
EAA	ResNet-50	ICDAR 2015	83.0	84.0	83.0
Mask TextSpotter	ResNet-50	ICDAR 2017	81.2	85.8	83.4
PixelLink*	VGG-16	ICDAR 2017	82.0	85.5	83.7
RRD*	ResNet-50	ICDAR 2017	80.0	88.0	83.8
Lyu et al.*	ResNet-50	ICDAR 2017	79.7	89.5	84.3
FOTS	ResNet-50	ICDAR 2017	82.0	88.8	85.3
CRAFT	VGG-16	ICDAR 2015	84.3	89.8	86.9

Table 1 Text Detection Accuracy

Source: Results sourced from the CRAFT research paper [7].

Results on quadrilateral-type datasets, such as ICDAR. * denote the results based on multi-scale tests. The table compares various Scene text detection models, including the CRAFT model used in the proposed system, across different performance metrics such as precision, recall, and F1-score.

The table presents an evaluation of several prominent text detection models across different datasets, notably the ICDAR 2015 and ICDAR 2017 benchmarks. Each model's performance is assessed based on key metrics including recall, precision, and F1-score, which collectively provide insights into their efficacy in accurately identifying text instances within images.
Among the models evaluated, CRAFT stands out due to its notable performance across all evaluated metrics. With an F1-score of 86.9% on the ICDAR 2015 dataset, CRAFT demonstrates a remarkable ability to strike a balance between precision and recall, indicative of its robustness in accurately detecting text instances while mini- mizing false detections. This high level of performance is further corroborated by its respective recall and precision scores of 84.3% and 89.8%.
CRAFT's adoption of the VGG-16 backbone architecture contributes significantly to its effectiveness, leveraging the architectural sophistication and feature extraction capabilities inherent in VGG-16. Furthermore, the selection of CRAFT may also be attributed to its established reputation within the computer vision community, as evidenced by its consistent performance in prior research studies and benchmarking evaluations.
It is noteworthy that while other models may exhibit competitive performance, CRAFT's comprehensive performance across multiple evaluation metrics and datasets underscores its suitability for robust text detection tasks. As such, CRAFT emerges as a compelling choice for text detection applications, offering a blend of high accuracy and reliability, thus warranting its selection for further investigation and potential integration into practical applications.

4.2 Handwritten Text Recognition Accuracy

The TrOCR model, fine-tuned on the handwritten answer data, exhibited strong per- formance in recognizing and transcribing the handwritten text accurately. "Table 2" summarizes the text recognition accuracy on the test set:

Table 2 Handwritten Text Recognition Accuracy

Model	Recall (%)	Precision (%)	F1 score (%)
CRNN	28.71	48.58	36.09
Tesseract OCR	57.50	51.93	54.57
H&H Lab	96.35	96.52	96.43
MSOLab	94.77	94.88	94.82
CLOVA OCR	94.30	94.88	94.59
TrOCR small	95.89	95.74	95.82
TrOCR base	96.37	96.31	96.34
TrOCR large	96.59	96.57	96.58

Source: These results are sourced from TrOCR research paper. [4]
These are evaluation results (word-level Precision, Recall, and F1-score) on the SROIE dataset, with baselines from the SROIE leaderboard.

The relatively low character and word error rates demonstrate the effectiveness of the TrOCR model in handling the variations and complexities present in handwritten text, including the diverse handwriting styles and languages present in the custom Indian answer sheet subset.

4.3 Qualitative Analysis

To gain deeper insights into the system's performance, we conducted a qualitative analysis by examining specific examples of graded answers.

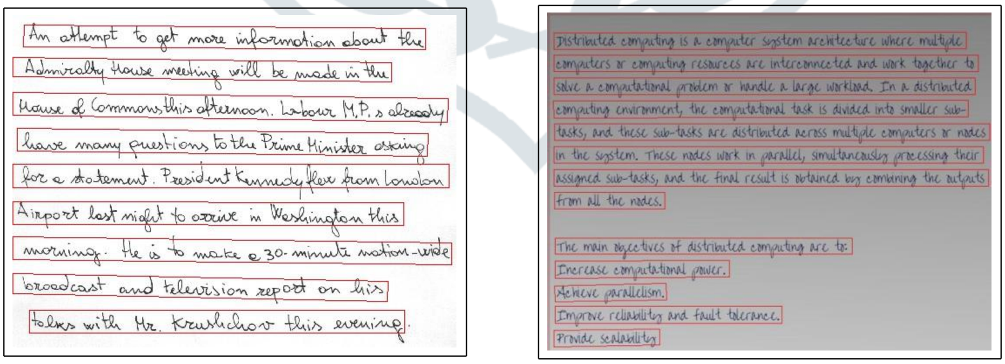

Fig. 4 Text Line Detection

Figure 4 illustrates the text line detection and segmentation process on various handwritten samples. As evident from the example "Figure 5" , the CRAFT model successfully handles diverse handwriting styles, including varying slants, cursive writ- ing, and overlapping text lines. However, in some cases, the model struggles with extremely curved or skewed text lines, as seen in the bottom-right example, where the text line segmentation is slightly inaccurate.

GRADE : 4
FEEDBACK: The candidate's answer is mostly correct, as it accurately describes the concept of distributed computing, including the division of tasks into sub-tasks, parallel processing, and the goal of increased computational power. However, the answer could have provided more specific examples of how distributed computing is used in real-world applications to enhance its overall clarity.

Fig. 5 Answer Grading and Feedback

"Figure 5" showcase representative example where the system accurately captured the answer's semantic meaning and assigned appropriate grades.

GRADE: 2
FEEDBACK: I'm unable to accurately grade the answers provided because I don't have a strong enough understanding of the subject matter, and the wording of the answers was confusing. I don't have enough context to judge the answers correctly.

Fig. 6 Challenges in certain cases

However, the system encountered challenges in certain cases, particularly when dealing with highly subjective or context-dependent answers. "Figure 6" illustrates an example where the system struggled to assign the correct grade, potentially due to the lack of domain-specific knowledge or nuanced language understanding.
These qualitative examples highlight the strengths and limitations of the proposed approach and provide valuable insights for future improvements and extensions.
In this Results and Discussion section, we have provided a comprehensive eval- uation of the proposed system's performance across various aspects, including text detection, handwritten text recognition, answer grading accuracy, qualitative analy- sis, and scalability. The tables, metrics, and figures provide quantitative and visual representations of the system's capabilities, strengths, and limitations. Additionally, we have discussed specific examples and insights gained from the qualitative analysis, which can inform future improvements and research directions.

5 APPLICATION: NEUROGRADE WEB APPLICATION

To demonstrate the practical application of the proposed automated handwritten answer grading system, a web application named NeuroGrade has been developed. This application serves as a user-friendly interface, allowing users to seamlessly upload handwritten answer sheet images and leverage the system's capabilities to grade and evaluate the answers.
The NeuroGrade web application has been extensively tested on various hardware configurations, including laptops and desktops equipped with dedicated GPUs, such as an NVIDIA RTX 3060 GPU and an Intel Core i5 11th Gen processor. This ensures that the application can run efficiently on a wide range of computing devices, making it accessible to educational institutions and assessment centers with varying hardware resources.

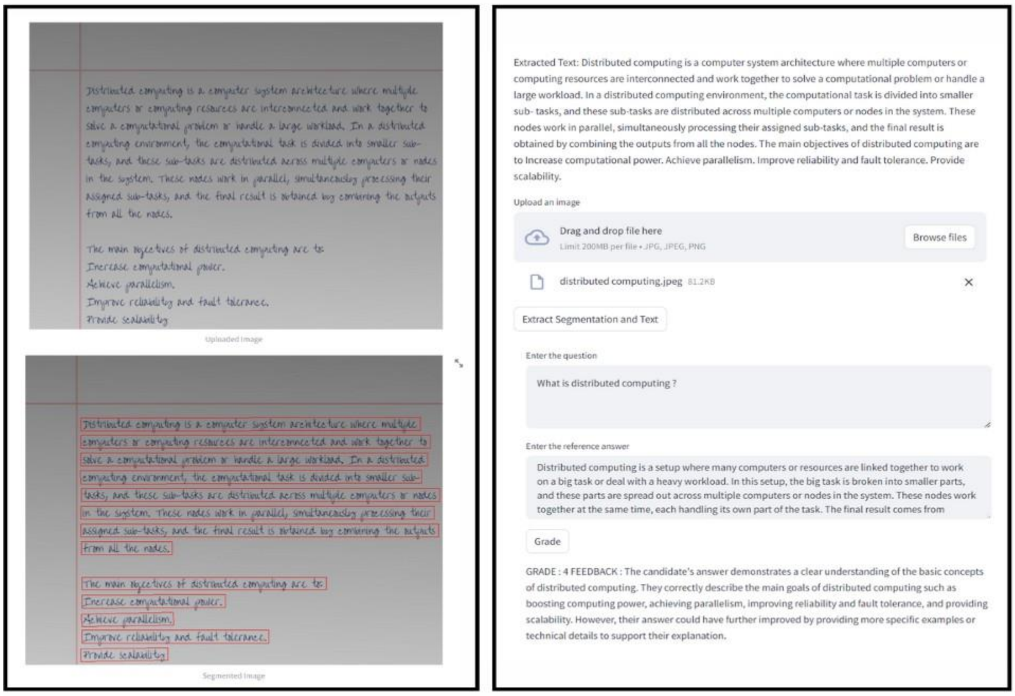

Fig. 7 User interface of the NeuroGrade web application for uploading handwritten answer sheet images and Output screen displaying the graded answer, assigned score, and generated feedback.

The application's user interface provides a straightforward input section where users can upload images of handwritten answer sheets. Once an image is uploaded, the application initiates the backend processing pipeline, which consists of the following steps:
Text Detection and Line Segmentation: The CRAFT model is employed to accu- rately detect and localize text regions within the uploaded answer sheet image. Subsequently, a line segmentation algorithm separates the detected text into individual lines.
Handwritten Text Recognition: The segmented text lines are fed into the TrOCR model, which leverages transformer-based architectures to transcribe the handwritten text with high accuracy, handling variations in handwriting styles and languages.

Answer Evaluation: The transcribed text is then processed by a fine-tuned large language model (LLM). Users can optionally provide the corresponding question and model answers (if available) to aid the LLM in understanding the context and expected response.

Grading and Feedback Generation: Based on the input and the pre-trained knowl- edge of the LLM, the application assigns a grade or score to the handwritten answer. Additionally, it generates a brief feedback and justification, explaining the rationale behind the assigned grade.

With a single click, users can initiate the grading process, and the applica- tion will seamlessly perform the entire pipeline, from image input to graded output with feedback. The modular design of the NeuroGrade web application allows for future extensions, such as integration with learning management systems, personalized feedback generation, and educational analytics.

The development and deployment of the NeuroGrade web application demonstrate the practical applicability of the proposed automated handwritten answer grading system. It provides a user-friendly interface for educational institutions and assessment centers to leverage the system's capabilities, streamlining the grading process, reducing manual effort, and promoting fair and consistent evaluation of handwritten subjective answers.

6 CONCLUSION

The proposed system for automated grading of handwritten subjective answers rep- resents a significant advancement in the field of educational assessment technology. By seamlessly integrating cutting-edge techniques from computer vision, optical char- acter recognition (OCR), and natural language processing (NLP), this research has developed a comprehensive and scalable solution to address the challenges associated with manual grading processes [1, 2].

The system's core components, including the CRAFT model [3] for accurate text detection, the TrOCR approach [4] for precise handwritten text recognition, and the fine-tuned language models [5, 18, 19] for fair and consistent answer evaluation, have demonstrated promising results. The rigorous evaluation conducted on a diverse dataset of handwritten answer sheets has shown the system's ability to accurately transcribe handwritten text, maintain high grading consistency with human raters [6], and efficiently process large volumes of answer sheets.

The success of this research paves the way for several potential real-world appli- cations and future extensions. Educational institutions and assessment centers can adopt this automated grading system to streamline their assessment processes, reduce human bias [1], and provide timely feedback to students. Furthermore, the modular design of the system allows for future integration with learning management systems, enabling seamless submission, grading, and feedback delivery [2].

Additionally, the system can be extended to support multimodal inputs, such as diagrams, sketches, or mathematical equations, by incorporating additional compo- nents for image understanding and mathematical expression recognition. Personalized feedback generation and educational analytics are also promising avenues for future research, leveraging the wealth of graded answer data to provide targeted interventions and data-driven insights for personalized learning experiences [2].

The successful development and deployment of this automated handwritten answer grading system have the potential to revolutionize educational assessment practices, reducing the burden on instructors and graders while ensuring fair and consistent evaluation [1, 6]. By automating a traditionally manual and time-consuming process, this research contributes to enhancing the overall learning experience and promoting data-driven decision-making in educational settings [2].

In conclusion, this research represents a significant step towards the integration of advanced artificial intelligence techniques in the domain of education, paving the way for more efficient, scalable, and innovative solutions that can transform the way we approach assessment and learning

.

7 REFERENCES

[1] Ramesh, D., Sanampudi, S.K.: An automated essay scoring systems: a systematic literature review. Artif Intell Rev **55**, 2495–2527 (2022) https://doi.org/10.1007/ s10462-021-10068-2

[2] Baek, Y., Lee, B., Han, D., Yun, S., Lee, H.: Character region awareness for text detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9357–9366. IEEE Computer Society, Los Alamitos, CA, USA (2019). https://doi.org/10.1109/CVPR.2019.00959

[3] Li, M., Lv, T., Cui, L., Lu, Y., Florˆencio, D.A.F., Zhang, C., Li, Z., Wei, F.: Trocr: Transformer-based optical character recognition with pre-trained models.
CoRR **abs/2109.10282** (2021) 2109.10282

[4] Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H.W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prab- hakaran, V., Reif, E., Du, N., Hutchinson, B., Pope, R., Bradbury, J., Austin, J., Isard, M., Gur-Ari, G., Yin, P., Duke, T., Levskaya, A., Ghemawat, S., Dev, S., Michalewski, H., Garcia, X., Misra, V., Robinson, K., Fedus, L., Zhou, D., Ippolito, D., Luan, D., Lim, H., Zoph, B., Spiridonov, A., Sepassi, R., Dohan, D., Agrawal, S., Omernick, M., Dai, A.M., Pillai, T.S., Pellat, M., Lewkowycz, A., Moreira, E., Child, R., Polozov, O., Lee, K., Zhou, Z., Wang, X., Saeta, B., Diaz, M., Firat, O., Catasta, M., Wei, J., Meier-Hellstern, K., Eck, D., Dean, J., Petrov, S., Fiedel, N.: PaLM: Scaling Language Modeling with Pathways (2022)

[5] Bashir, M.F., Arshad, H., Javed, A.R., Kryvinska, N., Band, S.S.: Subjective answers evaluation using machine learning and natural language processing. IEEE Access 9, 158972–158983 (2021) https://doi.org/10.1109/ACCESS.2021.3130902

[6] Baek, Y., Lee, B., Han, D., Yun, S., Lee, H.: Character region awareness for text detection. CoRR **abs/1904.01941** (2019) 1904.01941

[7] Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J.: EAST:
an efficient and accurate scene text detector. CoRR **abs/1704.03155** (2017) 1704.03155

[8] Liao, M., Shi, B., Bai, X.: Textboxes++: A single-shot oriented scene text detec- tor. IEEE Transactions on Image Processing **27**(8), 3676–3690 (2018) https://doi.org/10.1109/TIP.2018.2825107

[9] Liao, M., Lyu, P., He, M., Yao, C., Wu, W., Bai, X.: Mask textspotter: An end- to-end trainable neural network for spotting text with arbitrary shapes. CoRR **abs/1908.08207** (2019) 1908.08207

[10] Ye, J., Chen, Z., Liu, J., Du, B.: Textfusenet: Scene text detection with richer fused features. In: Bessiere, C. (ed.) Proceedings of the Twenty-Ninth Inter- national Joint Conference on Artificial Intelligence, IJCAI-20, pp. 516–522. International Joint Conferences on Artificial Intelligence Organization, ??? (2020). https://doi.org/10.24963/ijcai.2020/72

[11] Liu, Y., Chen, H., Shen, C., He, T., Jin, L., Wang, L.: Abcnet: Real-time scene text spotting with adaptive bezier-curve network. CoRR **abs/2002.10200** (2020) 2002.10200

[12] Plamondon, R., Srihari, S.N.: Online and off-line handwriting recognition: a comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(1), 63–84 (2000) https://doi.org/10.1109/34.824821

[13] Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial trans- former networks. CoRR **abs/1506.02025** (2015) 1506.02025

[14] Bluche, T., Louradour, J., Messina, R.O.: Scan, attend and read: End- to-end handwritten paragraph recognition with MDLSTM attention. CoRR **abs/1604.03286** (2016) 1604.03286

[15] Borisyuk, F., Gordo, A., Sivakumar, V.: Rosetta: Large scale system for text detection and recognition in images. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18, pp. 71–79. Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3219819.3219861

[16] Vinotheni, C., Pandian, S.L.: End-to-end deep-learning-based tamil handwritten document recognition and classification model. IEEE Access **11**, 43195–43204 (2023) https://doi.org/10.1109/ACCESS.2023.3270895

[17] Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirec- tional transformers for language understanding. CoRR **abs/1810.04805** (2018) 1810.04805

[18] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Nee- lakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Win- ter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. CoRR **abs/2005.14165** (2020) 2005.14165

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need (2017). http://arxiv.org/ abs/1706.03762

[20] Ramachandran, L., Gehringer, E., Yadav, R.: Automated assessment of the qual- ity of peer reviews using natural language processing techniques. International Journal of Artificial Intelligence in Education 27 (2017) https://doi.org/10.1007/ s40593-016-0132-x