IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 February 2025, accepted 7 April 2025, date of publication 17 April 2025, date of current version 5 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3561916

RESEARCH ARTICLE

Extracting Learning Data From Handwritten Notes: A New Approach to Educational Data Analysis Based on Image Segmentation and Generative AI

YUNYU ZHOU, CHENG TANG, (Member, IEEE), AND ATSUSHI SHIMADA, (Member, IEEE)
Faculty of Information Science and Technology, Kyushu University, Fukuoka 819-0395, Japan

Corresponding author: Atsushi Shimada (atsushi@ait.kyushu-u.ac.jp)

This work was supported in part by the Japan Science and Technology Agency (JST) Core Research for Evolutionary Science and Technology (CREST) under Grant JPMJCR22D1; and in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI, Japan, under Grant JP22H00551.

ABSTRACT While handwritten notes offer valuable insights into students' knowledge retention, traditional analysis methods are often time-consuming and limited in scope. This study introduces an efficient approach for educational data analysis by combining image segmentation with generative AI to extract learning insights from students' handwritten notes. Leveraging the Attention Multi-task U-Net for accurate segmentation and GPT-4o for content analysis, our method precisely identifies and categorizes text, charts, and formulas within notes. The extracted data provides educators with a detailed view of students' knowledge retention, highlights the areas students focus on, and identifies critical knowledge points that may be missing from notes. Our experiments on student notes from a Digital Signal Processing course demonstrate the method's high accuracy and significant efficiency improvements in teachers' review of student notes. This research contributes to educational technology and data mining by introducing an automated, scalable method that supports more personalized and effective educational strategies.

INDEX TERMS Attention multi-task U-Net, GPT-4o, handwritten notes, image segmentation, knowledge extraction, natural language processing.

I. INTRODUCTION

In the field of educational data analytics, transforming handwritten notes into meaningful learning data is gaining traction. Handwritten notes offer insights into student engagement and comprehension but analyzing them manually is labor-intensive and lacks scalability. In the context of modern education, automated note analysis methods are particularly important as the amount of educational data increases and the need for personalized learning grows. Handwritten notes contain a large amount of unstructured information, such as diagrams, formulas, and text, which can help teachers understand students' mastery of course content and what knowledge students may not be aware of. Therefore, it is important to study efficient and accurate handwritten note analysis methods for educational decision-making. In recent years, developments in image segmentation and natural language processing (NLP) have gradually addressed these issues; however, there are still major challenges in analyzing complex, unstructured handwritten note data.

Image segmentation techniques, such as U-Net and its variants, have demonstrated good results in many image segmentation tasks. However, many existing models still suffer from a lack of accuracy when dealing with complex layouts and diverse content structures in educational notes. For example, in educational environments, notes have different formats and content layouts, including dense mathematical symbols and hand-drawn graphs, which pose a significant challenge to traditional image segmentation models. Further improving the adaptability and accuracy of segmentation models, especially those that can simultaneously recognize elements such as text, formulas, and graphs, is crucial to improving

The associate editor coordinating the review of this manuscript and approving it for publication was Yin Zhang.

the accuracy of handwritten note analysis. Our study employs the Attention Multi-task U-Net model, which improves page segmentation accuracy by introducing an attention mechanism while using character segmentation as an auxiliary task to improve segmentation accuracy. The multi-task feature of this model enables it to handle multiple note elements, such as text, formulas, images, and tables, simultaneously, effectively compensating for the shortcomings of existing models in segmentation.

In recent years, the integration of attention mechanisms into segmentation models has shown substantial promise in focusing on key areas within complex data. For example, attention layers can help models prioritize dense regions of text or intricate diagrams, reducing the likelihood of misclassification and ensuring more precise segmentation. This targeted focus is especially beneficial for handling the intricate structures typical of handwritten notes, where text, images, and symbols are intermingled in non-standard formats. By leveraging these advancements, our study's model is designed to efficiently dissect multifaceted content, thus providing an effective foundation for subsequent analysis.

Similarly, advances in NLP technology have made it possible to efficiently understand the content of handwritten notes. However, traditional NLP models often have limitations in accurately interpreting unstructured handwritten text. The maturity of OCR technology has laid the foundation for the automatic recognition of handwritten text, and generative AI models, especially advanced models such as GPT-4o, have made significant improvements in semantic understanding. By deeply understanding and analyzing text content through generative AI, not only can key knowledge points be identified, but potential blind spots in student learning can also be anticipated, providing a more comprehensive view of student engagement with the material.

The research in this paper relies on the combination of OCR and generative AI, utilizing the advanced GPT-4o model, which accurately matches note content with textbook content through its powerful comprehension capabilities, identifying core knowledge points and discovering potential knowledge blind spots. GPT-4o's robust understanding of language and contextual nuances allows for a level of analysis that traditional NLP models struggle to achieve, particularly in recognizing contextual relevance between student notes and core curriculum concepts.

This combined approach of enhanced segmentation and generative NLP offers a scalable solution for extracting detailed learning insights from handwritten notes, enabling data-driven educational strategies that enhance teaching effectiveness. This method can not only be applied in actual teaching scenarios, enhancing instructors' ability to adapt lessons in real-time, but also provides new ideas for future educational data mining. As automated note analysis continues to evolve, the insights derived from student notes can form the basis for developing personalized educational feedback systems that adjust to each student's unique learning trajectory, ultimately contributing to adaptive and individualized learning environments. Through this approach, educators can more accurately monitor students' learning progress, identify specific areas that may require additional focus, and thus develop more targeted teaching strategies that foster improved educational outcomes.

In designing the model architecture, we prioritized computational efficiency and sensitivity to fine-grained features. After systematic comparisons with mainstream models (e.g., FCN, SegNet, SAM), the U-Net-based architecture was selected for its lightweight design and superior performance in capturing stroke-level details (achieving 0.9561 text region accuracy, 6.06% higher than SAM), as demonstrated in Section IV-B1. This choice aligns with the pipeline's requirement for rapid preprocessing while maintaining high segmentation precision.

II. RELATED WORK

A. APPLICATIONS OF IMAGE PROCESSING IN EDUCATION

1) SEGMENTATION AND CLASSIFICATION OF HANDWRITTEN NOTES

Current research has shown that image processing technologies can be effectively used to segment and classify handwritten notes, allowing for more efficient recognition of text, diagrams, and formulas. For example, Li et al [35] applied segmentation techniques to divide notes into text, mathematical expressions, and diagrams. This segmentation enables a more refined analysis of learning behaviors, offering insights into students' note-taking patterns and the effectiveness of their study methods. Further research has shown that models like Attention U-Net and U-Net++ are also can be applied in educational contexts, demonstrating high accuracy and multitasking capabilities in image data processing. However, the diversity and complexity of educational notes continue to challenge these techniques. This study optimizes segmentation models to address the complex demands of educational data analysis.

2) IMAGE SEGMENTATION TECHNOLOGIES IN EDUCATIONAL CONTEXTS

Image segmentation is vital for recognizing handwritten content in educational settings. Models like Attention Multi-task U-Net, known for their high precision and multitasking capability, have recently been widely applied in educational image data processing. These models accurately segment various parts of handwritten notes, such as text and graphs, which supports subsequent data analysis and enhances the overall understanding of students' learning progress.

3) APPLICATIONS OF DEEP LEARNING IN THE CLASSIFICATION OF HANDWRITTEN NOTES

Previous studies have shown a deep learning-based approach for classifying handwritten notes, significantly improving the organization and analytical efficiency of students' notes. Studies like these demonstrate how deep learning can enhance the classification and management of handwritten notes, contributing to the optimization of educational resources and the systematic organization of study materials.

B. APPLICATIONS OF NATURAL LANGUAGE PROCESSING IN EDUCATION

1) AUTOMATED EVALUATION AND ANALYSIS OF HANDWRITTEN CONTENT

Natural Language Processing (NLP) has been widely applied in the automated evaluation of handwritten notes. For example, Previous research in this area developed an automatic grading system that combines Information Retrieval (IR) and NLP technologies, achieving grading accuracy comparable to human evaluators. This system demonstrates the potential of NLP in the efficient evaluation of handwritten notes, reducing the workload on educators. Moreover, the combination of OCR and generative AI is gaining attention to recognize and understand complex handwritten content. This research further leverages the advanced generative AI model GPT-4o to enhance NLP's capability in understanding and processing complex handwritten data.

2) INTEGRATION OF OCR AND NLP TECHNOLOGIES FOR HANDWRITTEN TEXT RECOGNITION

For example, there are related studies introduced an automated grading system that integrates Optical Character Recognition (OCR) and NLP, significantly enhancing grading efficiency.

This integration allows for the automated recognition and comprehension of handwritten content, making it a feasible solution for large-scale educational applications.

3) APPLICATION OF GENERATIVE AI IN EDUCATION

Generative AI has shown promising potential in content analysis and evaluation within educational contexts. A new study employed GPT-4 to extract keywords and summarize answers from handwritten text, achieving accuracy comparable to manual grading.

This example illustrates the capability of advanced generative AI models to understand complex, unstructured handwritten content, highlighting their role in educational assessments.

C. THE POTENTIAL OF COMBINING IMAGE PROCESSING AND NLP IN HANDWRITTEN NOTE ANALYSIS

1) ADVANTAGES OF COMBINING IMAGE PROCESSING AND NLP

The integration of image segmentation and generative AI presents new possibilities for analyzing handwritten notes. Recent approaches leverage image segmentation, such as Attention Multi-task U-Net for page division, followed by generative AI models like GPT-4o for content analysis. This combined method enables efficient content recognition and provides valuable learning feedback, improving the efficiency and accuracy of educational data analysis. However, current methods still face challenges in educational contexts, such as the difficulty of recognizing complex handwritten content and handling unstructured data. Future research could further focus on designing advanced models tailored for large-scale educational data processing to enhance accuracy and operational efficiency in analyzing diverse handwritten content.

2) LIMITATIONS OF CURRENT METHODS AND FUTURE DIRECTIONS

Despite their effectiveness, existing image processing and NLP methods have limitations in educational contexts. For example, accurately recognizing complex handwritten scripts and handling unstructured data can be challenging. Future research can focus on more advanced models designed for large-scale educational data processing, which could enhance accuracy and operational efficiency in analyzing diverse handwritten content.

III. METHODOLOGY

A. METHOD OVERVIEW

In this study, we propose a novel method combining advanced image segmentation and generative AI techniques to analyze handwritten notes, aiming to capture detailed insights into students' learning behaviors. Our primary objective is to extract key information automatically and accurately from handwritten content, including text, graphs, and formulas, allowing educators to assess students' comprehension, knowledge retention, and areas of focus more efficiently. As educational technology progresses and the need for data mining grows, the demand for such automated methods has become increasingly urgent. Handwritten notes often contain a wealth of unstructured data, such as diagrams, formulas, and free text, making manual analysis costly. Automated methods enhance both the efficiency and accuracy of the analysis, and using AI to generate personalized feedback helps educators quickly identify knowledge gaps and learning needs.

Our approach leverages two main technologies: the Attention Multi-task U-Net model for precise image segmentation and the GPT-4o model for content understanding. The Attention Multi-task U-Net is utilized due to its high precision and ability to handle multiple tasks simultaneously, making it well-suited for segmenting diverse elements within student notes. This segmentation step is crucial for separating various content types, enabling a structured analysis of each note component. Following segmentation, the extracted content is analyzed by the GPT-4o model, which excels in processing unstructured handwritten data and generating coherent analyses. By integrating these technologies, our method not only achieves high accuracy in content identification but also enables in-depth analysis of students' learning patterns through automated data extraction.

This methodology represents a significant step forward in the educational data mining field, offering a scalable, data-driven approach to understanding student engagement and knowledge acquisition.

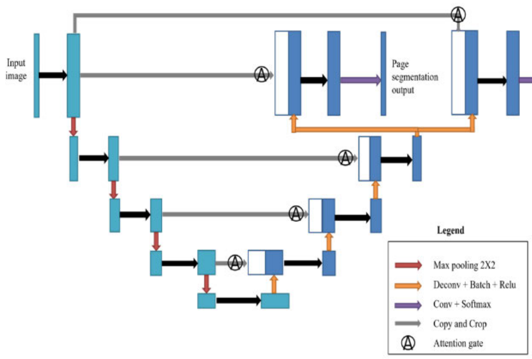

FIGURE 1. Attention multi-task U-net.

B. ATTENTION MULTI-TASK U-NET [6]

The Attention Multi-Task U-Net used in this study is designed to segment various elements in handwritten notes, such as text, charts, and formulas. As shown in Fig.1, by incorporating attention and multi-task mechanisms into the U-Net architecture, the model achieves greater segmentation accuracy and efficiency, which is essential for the complex layouts often found in handwritten notes.

During the image preprocessing phase, this study first uses image enhancement techniques, such as contrast enhancement and noise reduction, to improve the clarity of handwritten content. This ensures that the model can efficiently segment image content even in complex backgrounds. Subsequently, The Attention Multi-Task U-Net segments preprocessed images into distinct regions, including text, diagrams, and formulas. This segmentation provides structured data for subsequent analysis and enhances model accuracy by isolating key content types.

The process begins with pre-processing the input images to enhance contrast and reduce noise, which improves the clarity of the elements within the notes. Following this, the Attention Multi-task U-Net model performs segmentation by dividing the page into distinct regions.

The attention mechanism incorporated in this model enables it to focus on relevant parts of the image, enhancing its ability to differentiate between textual and non-textual elements. This capability is especially beneficial in educational settings, where handwritten notes contain a mix of diverse content types.

The attention mechanism is added to the U-Net to enable the model to focus selectively on important areas within the notes. Traditional U-Net models distribute focus evenly across the image, which can reduce performance when processing complex and mixed-content documents. The attention layer allows the model to identify and concentrate on relevant features, such as dense text or intricate diagrams, improving precision by filtering out less critical areas. This focused processing is particularly beneficial for segmenting student notes, where small details can carry significant meaning.

The multi-task learning feature allows the model to handle multiple segmentation tasks simultaneously, such as distinguishing between text, tables, and formulas. By training the model to handle these tasks simultaneously, multitask learning enhances generalization to different content types, leading to better segmentation of various elements in educational notes. In handwritten note analysis, common content types such as text, formulas and diagrams often alternate on the same page. Multi-task learning integrates the character segmentation and page segmentation tasks into a single segmentation process, avoiding the repetition of separate processing steps in previous methods and thus greatly increasing processing speed. On the other hand, character segmentation as a side task also improves the segmentation accuracy of the page segmentation task.

Combining attention and multi-tasking in the U-Net architecture enables the segmentation model to achieve high accuracy in a single, integrated process, ensuring that the extracted elements can be effectively used in subsequent NLP and generative AI analysis.

As shown in Fig.2, this is an example of page segmentation for handwritten notes with text, formulas, graphs, and tables.

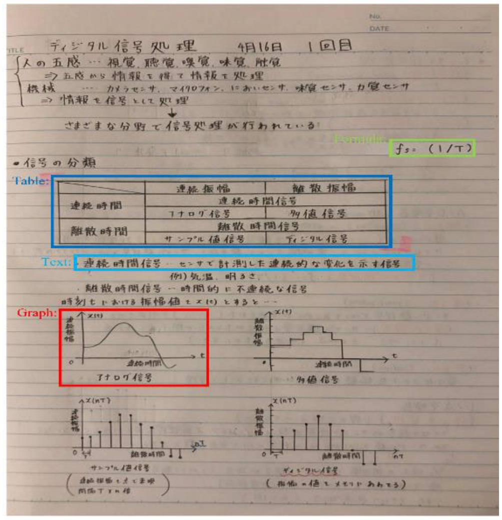

FIGURE 2. Handwritten note page segmentation example.

C. GPT-4O

Following segmentation, this study employs GPT-4o, an advanced generative AI model, to analyze the content extracted from student handwritten notes. The purpose of GPT-4o in this process is to understand and interpret text, diagrams, and formulas. The results of the segmentation of students' notes are matched with the content of the textbook to search for the keywords that are of most concern to the students, and the knowledge points that the students missed. OCR technology is first applied to the segmented text regions to convert handwritten text images into machine-readable text, improving accuracy in subsequent processing. Additionally, by constructing a knowledge base, GPT-4o can precisely match text with specific knowledge points, generating targeted learning feedback.

To prepare the content for analysis, OCR (Optical Character Recognition) is first applied to the segmented text regions to convert images of handwritten words into machine-readable text. OCR is essential in this stage because it bridges the gap between handwritten content and the NLP model, enabling accurate text extraction from various handwriting styles. By using OCR in conjunction with image segmentation, we ensure that only well-defined and legible text enters the NLP pipeline, improving processing accuracy.

The GPT-4o model then processes the OCR-transformed text and applies its deep language understanding capabilities to analyze the content. For example, the model can extract core concepts, identify areas of repeated emphasis, and even infer knowledge gaps based on note content. Additionally, GPT-4o can generate feedback to help teachers identify students' learning challenges and weaknesses, enabling the design of personalized teaching plans.

One of the strengths of GPT-4o is its capacity to handle complex, unstructured text, which is often found in educational notes. Unlike traditional NLP models, GPT-4o can discern context, extract key points, and generate summaries, making it highly suitable for interpreting student notes. For example, the model can highlight core concepts, pinpoint areas of repeated emphasis, and even infer gaps in knowledge based on the content covered in the notes. Additionally, GPT-4o's generative capacity allows it to formulate potential feedback that educators can use to guide students effectively.

This integration of OCR with GPT-4o provides a streamlined approach to content analysis, transforming handwritten notes into structured data insights. The outputs generated from GPT-4o, such as extracted keywords, summaries, and inferred understanding levels, serve as valuable tools for educators to assess and tailor their teaching methods to better align with students' learning needs.

D. INTEGRATED FRAMEWORK

As shown in Fig.3 and Fig.4. The integrated framework combines image segmentation and generative AI to create a streamlined process for analyzing handwritten notes. This framework is designed to facilitate the automated extraction, processing, and interpretation of student notes, making it possible to gather detailed insights into learning behaviors.

Overall process: the workflow of the framework starts with the input of handwritten notes, which are subsequently processed by an image segmentation model (Attention Multi-task U-Net) to segment the different content categories. After segmentation is complete, the OCR technology will convert the text area (image format) and save it in text format (.txt). Although the current GPT model has multimodal capabilities, our experiments revealed that GPT is unable to effectively recognize Japanese text in pictures, particularly for handwritten notes, which are often written in a non-standardized manner. Considering this, we developed the OCR module for the extraction of text from segmented text and table regions. This enables the Japanese text to be entered directly into the GPT, obviating the necessity for the GPT to recognize the image once more. This results in enhanced efficiency for the GPT and improved matching accuracy. The images and formulas are saved in image formats for subsequent processing. Ultimately, the GPT-4o model, as shown in the Fig.5, performs a matching operation on the data to generate the matching results between student notes and textbook content. The modules within the framework engage in a structured sequence of interactions. For example, the image segmentation module outputs the segmentation results of text, formulas, graphs, and tables. The text within the segmented table is extracted by the OCR module and saved as a text file with the .txt file. Afterwards, the textual data are input into the GPT-4o. This interaction creates a data processing framework that ensures processing efficiency and minimizes manual intervention.

Data storage and processing mechanism: all categories of data extracted from the textbook as shown in the Fig.4, the. txt files and image files are saved into our new constructed knowledge base of GPT-4o. Through this knowledge base, when dealing with each input of handwritten note data, GPT can directly match the search in the knowledge base, which improves the efficiency of GPT greatly. By integrating an image segmentation model, OCR, and GPT within a framework, this approach provides an efficient and accurate solution for extracting deep data from handwritten notes.

This framework design not only automates the data extraction but also prepares it for further analysis. It helps in the application of personalized learning feedback and adaptive educational recommendations.

E. INNOVATION AND ADVANTAGES

1) ATTENTION MECHANISMS AND THE INNOVATIVE COMBINATION OF MULTI-TASK LEARNING

An important innovation of this method is the introduction of the attention mechanism and multitask learning into the U-Net model. The attention mechanism enables the model to segment handwritten notes and achieve high segmentation accuracy even in complex datasets. Multi-tasking further improves the model's page segmentation accuracy, and in our previous experiments we found [6] that the character segmentation task will greatly improve the accuracy of the page segmentation task. This combination significantly enhances the efficiency and accuracy of extracting data from a complex layout dataset such as handwritten notes.

2) GPT-4O FOR DEEP CONTEXT UNDERSTANDING

Another important innovation is the deep semantic analysis of segmented content using the GPT-4o model. To match handwritten notes with textbook content for searching, we need the model can understand the meanings of different textual expressions and match them correctly for the operation. GPT, on the other hand, is one of the best-performing large models for semantic understanding, so we chose this model for the NLP part of the function to output the matching results and

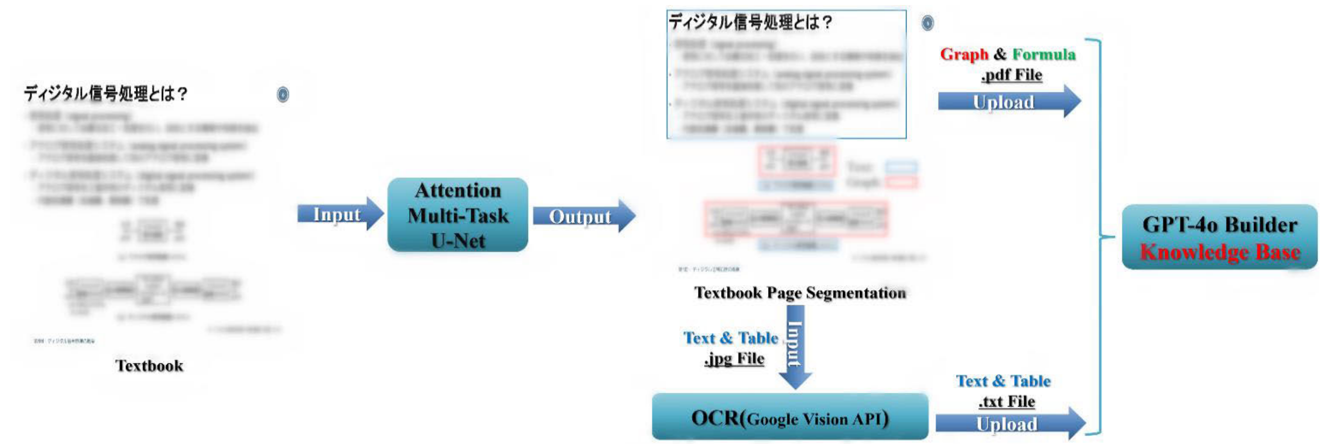

FIGURE 3. Process of building a knowledge base of textbooks.

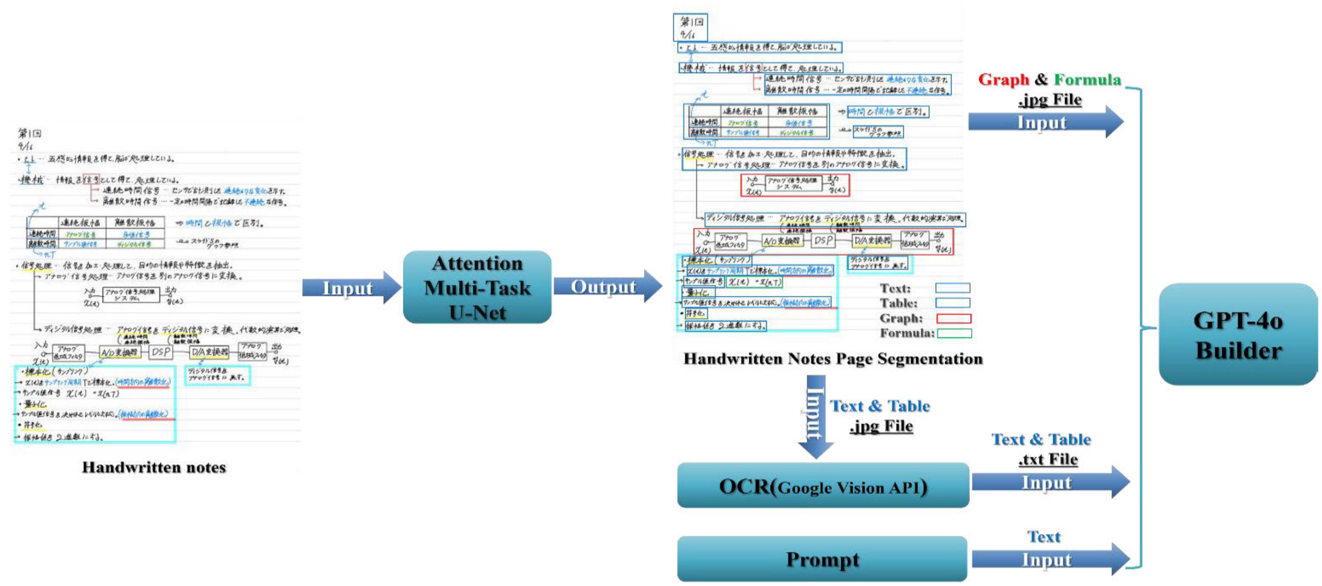

FIGURE 4. Process for student handwritten notes.

provide educators with in-depth insights from students about their learning situation efficiently and accurately.

3) INTEGRATION OF IMAGE SEGMENTATION MODELS, OCR AND NLP MODELS

The method combines image segmentation, OCR and NLP models within a framework that provides a concise framework for handwritten notes data extraction. By integrating these processes within a single framework, the method minimizes human intervention, reduces errors that may result from human manipulation, and increases the efficiency of analyzing student handwritten note data. This integration makes it possible to process large amounts of notes quickly and has great potential to be applied to diverse educational scenarios.

4) ADVANTAGES OVER TRADITIONAL METHODS:

Compared to traditional methods that typically rely on manual review, this approach is highly automated and reduces the time and effort required by instructors to assess student notes. The integrated framework provides content-matching results between students‘ handwritten notes and the textbook, enabling educators to better understand students’ engagement and mastery of course content. In addition, the method’s ability to highlight key concepts from the textbook and identify potential gaps in student knowledge supports personalized learning strategies and has the potential to become an adaptive educational tool.

5) FUTURE POTENTIAL

The methodology’s scalability and flexibility make it ideally suited for future growth. As the field of education increasingly embraces digital tools, this method can be adapted to a variety of learning environments, from classrooms to online courses. Its ability to analyze a wide range of handwritten data, coupled with its adaptable framework, makes it valuable for applications in educational data mining and

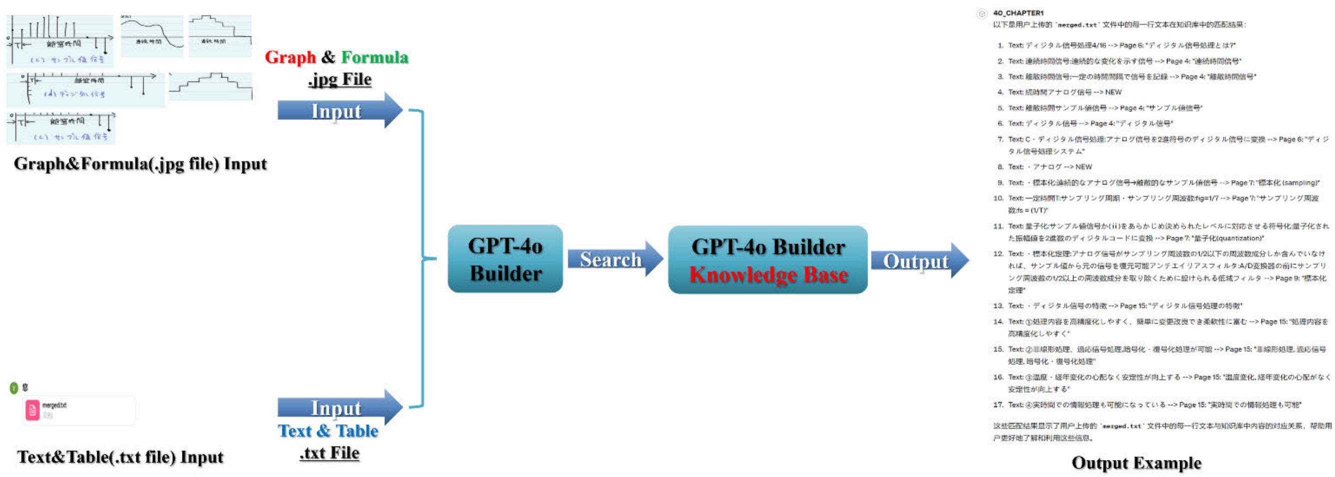

FIGURE 5. Matching result example.

adaptive learning, and has great potential for future research and practice.

IV. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL DESIGN

This section describes the overall design of the experiment to evaluate the segmentation accuracy, content matching accuracy of the proposed method. The experimental data are handwritten notes from students 2021 and 2022 of Signal and System Processing.

Data preparation: Student handwritten notes were obtained from students' Digital Signal Processing courses to ensure data integrity. The experiment began with data preprocessing, where all note images were standardized to the same resolution to mitigate the impact of size variations. Furthermore, to enhance the robustness of the model across diverse writing styles, we selected samples that included various handwriting styles, layout designs, and content complexities.

B. EXPERIMENT RESULT

1) SELECTION OF MODELS

Given the limited dataset scale (2,000 annotated images) in this study, we deliberately excluded computationally intensive models like DeepLabV3+ during the architecture selection phase. The primary rationale for adopting U-Net as our lightweight architecture stems from its exceptional sensitivity to fine-grained visual features - a critical requirement for accurately segmenting handwritten student notes. The characteristic skip connections in U-Net effectively preserve spatial details through multi-scale feature fusion, which proves particularly advantageous for capturing intricate stroke patterns in free-form handwriting.

Furthermore, the computational efficiency of U-Net aligns with our pipeline design philosophy where segmentation serves as an intermediate processing step rather than the terminal objective. As shown in Table 1. Our quantitative comparison with FCN and SegNet demonstrates U-Net's superior

TABLE 1. Comparison of U-Net accuracy with other neural networks.

	Total Ac/F1	Back ground	Text	Formula	Graph
FCN	0.7526/ 0.7135	0.8436	0.7558	0.6424	0.5017
SegNet	0.7828 /0.7562	0.8706	0.7927	0.7154	0.5983
SAM	0.8926 /0.8753	0.9326	0.9015	0.8571	**0.8916**
Original U-Net	0.8162/ 0.7869	0.9137	0.8253	0.7346	0.5891
Attention Multi Task U-Net	**0.9425/ 0.9264**	**0.9733**	**0.9561**	**0.8651**	0.8263

performance in text (0.8253) and formula (0.7346) regions. Although SegNet shows marginal advantage (0.5983 vs 0.5891) in graphic element segmentation, U-Net's balanced performance across all categories and superior computational efficiency make it the optimal choice for our educational data analysis task.

Furthermore, this study conducts systematic comparative analyses with mainstream segmentation models. As shown in Table 1, while the baseline SAM model outperforms the traditional U-Net architecture in overall performance (e.g., Total region F1-score 0.8753 vs. 0.7869), the proposed Attention Multi-Task U-Net demonstrates significant advantages in critical metrics: achieving a text region segmentation accuracy of 0.9561 (6.06% higher than SAM) and marginally leading SAM in formula regions (0.8651 vs. 0.8571). Notably, SAM excels in graphic regions (0.8916 vs. 0.8263). However, considering that text elements are much more prevalent than graphic elements in student handwritten notes. and considering the segmentation module's role as a preprocessing step in the multimodal analysis pipeline, the efficiency gains from lightweight design and

TABLE 2. Comparison of accuracy rates by region.

	Total Ac/F1	Back ground	Text	Formula	Graph
Original U-Net	0.8162/ 0.7869	0.9137	0.8253	0.7346	0.5891
Multi-Task U-Net	0.8728 /0.8475	0.9406	0.8927	0.7854	0.6083
Attention Multi Task U-Net	0.9425/ 0.9264	0.9733	0.9561	**0.8651**	0.8263
Attention R2 Multi-Task U-Net	**0.9458/ 0.9351**	**0.9782**	**0.9608**	0.8647	**0.8315**

TABLE 3. Matching accuracy comparison.

Chapter	Gemini	Claude	GPT-4o
1	0.908	0.904	0.938
2	0.913	0.909	0.941
3	0.890	0.883	0.916
4	0.888	0.882	0.914
5	0.901	0.896	0.925
6	0.897	0.891	0.926

domain adaptability become decisive factors. The experimental results indicate that Attention Multi-Task U-Net offers superior engineering applicability in precision-efficiency trade-offs. Particularly in resource-constrained deployment scenarios, its modular design enables seamless integration with downstream LLM components, providing crucial technical support for building end-to-end intelligent note analysis systems.

2) IMAGE SEGMENTATION MODEL (ATTENTION MULTI-TASK U-NET)

The main content of (subsection I) is mainly quoted from our paper "Improvement of image segmentation model for handwritten notebook analytics" presented at ICIP 2023.

Model training: The dataset we used in our experiments is the handwritten notes produced by students in the Digital Signal Processing course between 2021 and 2022. There are a total of 1980 images. We split it into 70% training set, 15% validation set, and 15% test set. The accuracy results shown in this paper are from the model's performance on the test set.

Evaluation metrics: segmentation accuracy and F1 score are used to assess the overall performance of the model. We calculate the accuracy of the model in each region by Equation 1. C represents the number of pixels that are correctly segmented, and L represents the number of pixels in the label.

$$\text{Accuracy} = C / L \tag{1}$$

To verify the performance of the model, we did some comparison experiments with other U-Net. Comparison of the multi-task U-Net model and the original U-Net: As shown in Table 2, it can be clearly found that the multi-task U-Net improves the accuracy over the original single task U-Net in the text(6.74%) and formula(5.08%) region and improves 1.92% in the graph region.

This indicates that in the multi-task model, the features learned by the character segmentation task assist the model in correctly identifying text and formula regions in the page segmentation task. Table 2 clearly demonstrates the advantages of incorporating the attention mechanism in the Attention Multi-task U-Net model, with significant improvements across regions. Notably, accuracy in the graph region increased by 21.8% compared to the multi-task U-Net model, while the text and formula regions showed respective improvements of 6.34% and 7.97%. These gains highlight the effectiveness of the attention mechanism in enabling the model to focus on specific features, particularly in regions with dense graphical elements and complex structures. This improved precision across multiple regions underscores the model's adaptability and accuracy in handling varied and complex note layouts.

As shown in Table 2, the progression from the original U-Net to the Attention Multi-task U-Net illustrates a steady improvement across all evaluated regions. The table highlights that while the multi-task U-Net model provided modest gains, the attention-enhanced version of the model made substantial advancements in handling graph regions with dense content and distinguishing between various page components. This comparison affirms the enhanced segmentation capabilities of the Attention Multi-task U-Net, positioning it as a robust model for processing complex handwritten educational notes.

3) CONTENT MATCHING ACCURACY OF GPT-4O

GPT-4o demonstrates strong content matching accuracy, effectively aligning handwritten notes with textbook content. This matching capability enables GPT-4o to accurately identify and associate key concepts from students' notes with corresponding curriculum materials, providing insights into student understanding and highlighting potential knowledge gaps. Such functionality is invaluable in automated educational assessments, where accurately comparing student-generated content to course material is essential for personalized feedback.

In selecting the most suitable model, we conducted a preliminary comparison of GPT-4o, Gemini, and Claude, evaluating each model's accuracy in matching complex and technical content within handwritten notes. As shown in Table 3, GPT-4o consistently outperformed the other models, particularly in handling challenging terminology and technical expressions. This superior accuracy allowed GPT-4o to achieve higher alignment rates with textbook content, making it the preferred choice for the NLP function in this study.

TABLE 4. GPT-4 accuracy of chapter 1-11.

Chapter	Text & Table	Graph	Formula	Corresponds
1	0.850	0.891	0.943	0.861
2	0.849	0.919	0.914	0.866
3	0.823	0.895	0.944	0.849
4	0.818	0.908	0.914	0.842
5	0.829	0.890	0.949	0.856
6	0.832	0.893	0.921	0.853
:7	0.844	0.881	0.946	0.865
8	0.836	0.917	0.931	0.862
9	0.869	0.883	0.932	0.879
10	0.816	0.896	0.915	0.837
11	0.827	0.901	0.924	0.834

TABLE 5. GPT-4o accuracy of chapter 1-11.

Chapter	Text & Table	Graph	Formula	Corresponds
1	0.939	0.901	0.958	0.938
2	0.945	0.937	0.933	0.941
3	0.917	0.912	0.952	0.916
4	0.912	0.921	0.925	0.914
5	0.923	0.908	0.961	0.925
6	0.919	0.913	0.936	0.926
7	0.936	0.894	0.957	0.937
8	0.931	0.929	0.946	0.932
9	0.953	0.902	0.947	0.945
10	0.908	0.912	0.923	0.913
11	0.912	0.916	0.935	0.914

The experiment analyzed handwritten notes from 145 students in a Digital Signal Processing course, covering 11 chapters of content. We observed that GPT-4o's accuracy in processing text improved by approximately 10% compared to its predecessor, GPT-4. This improvement was largely achieved through enhanced prompt engineering, which optimized the model's capacity to interpret specific technical terms accurately. Furthermore, Tables 4 and 5 provide a detailed breakdown of GPT-4o's accuracy across chapters and content types, illustrating its performance consistency. These results validate GPT-4o's applicability in educational contexts, making it a robust tool for automated analysis and feedback generation. The model's ability to align content with precision has significant implications for future applications in educational data analysis and adaptive learning systems.

$$Accuracy = \frac{correct\ classifications}{all\ classifications} \tag{2}$$

4) FURTHER ANALYSIS OF CONTENT MATCHING

By analyzing the results of content matching, it is possible to generate in-depth insights into student learning patterns and areas of focus. As shown in Table 6, GPT-4o was used to extract a list of keywords from the textbook content of Chapter 1, with these keywords presented in the first column. Following this, the matching results allowed us to count how many A-level students recorded each keyword. Observing the table, it is immediately apparent that the keywords most frequently recorded by students are "A/D 変換器" and "D/A 変換器," indicating a high level of attention toward these concepts.

TABLE 6. Chapter 1 keywords.

Keyword	Textbook Content	2021	2022
ディジタル信号処理とは?	信号処理 (signal processing) …	35	21
ディジタル化の利点	雑音に強い 劣化しない…	49	38
ディジタル信号処理の特徴	数式通りの処理内容を...	47	37
信号処理システムの時不変性	ある信号x(t)が時間だけ遅れた入力…	37	29
信号処理システムの因果性	システムへ信号が入力された後…	39	26
信号処理システムの線形性	線形性 ある変換{ }を行うシステ…	31	25
A/D 変換器	標本化 (sampling) 連続的なアナロ…	**53**	**43**
D/A 変換器	2進符号のディジタル信号をアナログ信号に戻す …	**54**	**41**
信号の形態	時刻における信号の振幅値をx(t)と表現 4種類…	48	40
情報と信号	ヒトの五感 視覚,聴覚,嗅覚,味覚,触覚 ヒトは…	42	35
着目点	時間成分の連続性 線形性 時不変性 因果…	44	38

Note: The Textbook Content was too long to be fully displayed in the table, therefore the use of ellipses (…) was used in the table to replace the textbook content that could not be displayed.

Although teachers often have structured teaching plans that highlight the relative importance of various content areas, it is challenging for them to gauge precisely which topics capture their students' focus. By using content matching analysis, educators can quickly identify what students are concentrating on. This information allows teachers to tailor their instruction more effectively, focusing on areas where student engagement is strong or addressing topics where comprehension may be lacking. For students, this method provides a valuable tool for self-assessment, helping them to review their notes and identify key knowledge points that may have been overlooked. This proactive review can facilitate deeper comprehension and retention of essential course content.

Table 6 's data allows us to explore patterns in students' understanding and note-taking behavior. Through analysis,

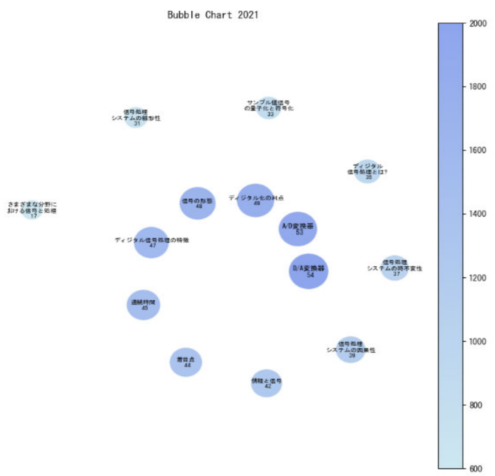

FIGURE 6. 2021 keywords bubble chart.

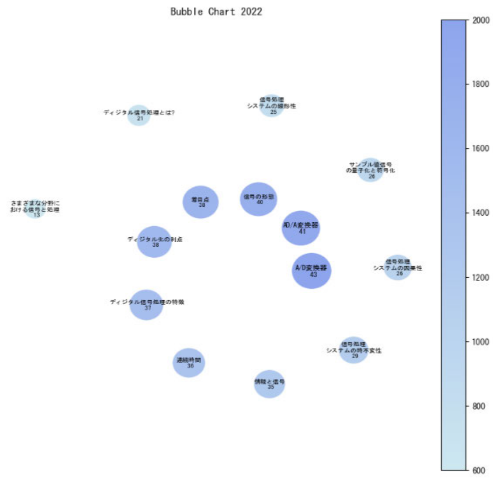

FIGURE 7. 2022 keywords bubble chart.

educators can assess the breadth and depth of students' focus on different topics, thus gaining insights into where students might require additional support. For instance, a high focus on "A/D 変換器" suggests that this concept is either challenging or particularly engaging for students, signaling an area where further instructional emphasis could be beneficial. On the other hand, keywords that receive minimal attention may reflect topics students find either easier or less relevant, guiding teachers in curriculum adjustments.

Using the data from Table 5, we generated bubble charts, shown in Figures 6 and 7, to illustrate student attention to these keywords across the 2021 and 2022 classes. The bubble chart visually presents keyword attention in a circular pattern, making it easy to compare the focus on each concept across different groups. Specifically, Fig. 6 displays data for the 2021 class, while Fig. 7 represents data from the 2022 class. This year-over-year comparison offers educators valuable insights into shifting patterns of student interest and understanding over time. For example, if certain keywords show a declining trend in attention, it may indicate the need to reinforce these topics in future sessions.

Note: The Textbook Content was too long to be fully displayed in the table, therefore the use of ellipses (...) was used in the table to replace the textbook content that could not be displayed.

Furthermore, the bubble charts enable educators to quickly observe how attention to specific knowledge points changes among different student groups. This visualization can reveal trends that help educators adapt their teaching to evolving student needs. For instance, if attention to a particular concept, like "A/D 変換器," increases significantly in one cohort compared to another, this may suggest that the curriculum or teaching approach has impacted student engagement with that topic. Such insights are useful for continuous improvement in curriculum design, enabling more focused, responsive teaching practices.

For students, these visual representations are equally beneficial. The charts provide a way to benchmark their understanding against their peers' collective focus, helping them identify topics they may have under-emphasized. By aligning their study efforts with areas that receive substantial attention from their classmates, students can reinforce learning on critical topics, fostering a well-rounded comprehension of the material. Additionally, understanding the focus trends of past student groups can help them anticipate which areas might require more attention and better prepare for exams or assessments.

In conclusion, content matching analysis and its visualization through bubble charts serve as powerful tools for enhancing educational outcomes. For teachers, they provide clear indicators of which topics may need more attention, adjustment, or support in the classroom. For students, they offer a structured method to identify and concentrate on key learning areas, promoting a more effective, data-driven approach to study and review. This dual benefit underscores the value of content matching in both guiding instructional strategies and supporting student learning practices.

V. DISCUSSION

A. KEY FINDINGS

This study proposes an efficient method for analyzing handwritten note data by combining image segmentation and generative AI models. Experimental results show that the integration of the Attention Multi-task U-Net model with GPT-4o offers significant advantages in segmentation accuracy and content understanding, enabling complex handwritten notes to be processed and analyzed in an automated manner. This combination allows for the accurate extraction of student learning data, the identification of key concepts, and the detection of knowledge blind spots, thereby providing educators with a quantitative approach to analyze student learning behaviors.

One of the main advantages of this method is its ability to handle complex layouts and diverse handwriting styles, which are common in student notes. The Attention Multi-task

U-Net model performs well in distinguishing between text, formulas, and graphical elements, thanks to the multi-tasking capabilities and attention mechanism that focus on specific features within each content type. This results in higher segmentation precision, even for intricate elements like diagrams and mixed-content layouts. The GPT-4o model complements this by accurately interpreting and linking segmented content to textbook material, which aids in identifying the alignment of students' notes with the curriculum. Such precision is crucial for providing teachers with insights into student comprehension and areas that might need further emphasis.

Compared to traditional manual methods, which are labor-intensive and time-consuming, this automated approach enhances both accuracy and efficiency. Educators can save significant time by using this method, allowing them to focus on instructional improvements based on data-driven insights. Furthermore, the method provides students with targeted feedback on their understanding and knowledge gaps, helping them better prepare for assessments. In summary, this integrated approach not only offers scalability and reliability but also holds potential for broader applications in educational data analysis and personalized learning support.

B. ADVANTAGES OF THE METHOD

The main advantages of this method include:

Automation and Efficiency: By integrating image segmentation, OCR, and NLP, the method achieves automated analysis of handwritten notes, greatly improving processing efficiency and making it suitable for large-scale applications. Unlike manual methods, which require significant time and effort, this approach enables educators to quickly process and analyze large volumes of notes, saving valuable time and resources. This automation is especially beneficial in educational environments where student data is continuously generated, allowing for more responsive and timely analysis.

Diverse Data Processing: The combination of Attention Multi-task U-Net and GPT-4o models allows the method to handle multiple content types, including text, charts, and formulas. This enhances the comprehensiveness of content analysis, as it accurately segments and interprets various elements that may appear in educational notes. The model's attention mechanism specifically aids in distinguishing between different data types within a single page, ensuring that each content type is analyzed accurately. This capability allows the method to adapt to various subjects and note structures, making it broadly applicable across educational disciplines.

Data Insights: The integration of generative AI provides deep insights into learning behaviors, allowing educators to identify patterns in students' study habits and knowledge gaps. By matching handwritten content with textbook material, the model enables a detailed understanding of which topics students are focusing on, what concepts they might be struggling with, and where they may need additional support. This data-driven insight is invaluable for educators seeking to provide targeted feedback and personalized guidance to students.

Lightweight model components: The selection of the Attention Multi-task U-Net over computationally intensive alternatives (e.g., SAM) proved critical to balancing accuracy and efficiency. As shown in Table 1, the model's lightweight architecture achieved a text segmentation F1-score of 0.9264 making it ideal for integration with downstream NLP components. This design choice underscores the importance of task-specific model optimization in educational data analysis pipelines.

However, the method has some limitations. For instance, OCR and NLP models may encounter difficulties with complex or poorly legible handwriting, potentially leading to errors in text recognition and interpretation. The model may also need optimization to handle different writing styles more effectively, particularly in cases where students use unconventional characters or symbols. Addressing these limitations would involve fine-tuning OCR algorithms and enhancing model adaptability, ensuring robust performance across diverse handwriting patterns.

C. COMPARISON WITH EXISTING RESEARCH

This study's integrated approach significantly enhances the effectiveness of traditional image processing and natural language processing (NLP) in educational data analysis. By combining multiple techniques, it not only improves analysis efficiency but also enhances content recognition accuracy. In contrast to existing methods based on manual review or single segmentation models, this study offers a more efficient and automated solution for analyzing handwritten notes.

Previous research has explored various methods for processing handwritten documents. For instance, Jo et al. [48] proposed a method that separates handwritten and machine-printed components in documents using convolutional neural networks (CNNs). Their approach focuses on pixel-level classification to handle overlapping components, demonstrating the potential of CNNs in document analysis. However, their method primarily addresses the separation of handwritten and printed text without delving into the semantic understanding of the content.

In another study, introduced a novel approach to segment handwritten and printed text, particularly in scenarios where they overlap. They developed a dataset, SignaTR6K, and a model architecture that outperforms prior work in segmentation tasks. While their work advances segmentation techniques, it does not integrate NLP components for content analysis.

Furthermore, such as the use of generative AI models, such as GPT-4, to extract keywords and generate summaries from handwritten text, achieving accuracy comparable to manual grading. This highlights the potential of generative AI in understanding handwritten content. However, their focus is primarily on content generation rather than the integration of image segmentation and NLP for comprehensive analysis.

In contrast, our study integrates advanced image segmentation models, specifically the Attention Multi-task U-Net, with generative AI models like GPT-4o. This combination allows for precise segmentation of handwritten notes into various components, such as text, formulas, and diagrams, followed by semantic analysis to extract meaningful insights. By leveraging the strengths of both image processing and NLP, our approach provides a holistic solution for educational data analysis, enabling automated extraction of student learning data, identification of key concepts, and detection of knowledge gaps.

Compared to single-model studies, our method's integration of multiple techniques enhances both efficiency and accuracy. The Attention Multi-task U-Net model improves segmentation accuracy by focusing on specific features within complex page layouts, while GPT-4o's advanced language understanding capabilities facilitate accurate content matching and analysis. This integrated approach addresses the limitations of previous methods by providing a comprehensive tool for analyzing handwritten notes, supporting educators in quantitatively assessing student learning behaviors and tailoring instructional strategies accordingly.

In summary, while existing research has made significant strides in either image segmentation or NLP for handwritten document analysis, our study bridges the gap between these domains. By combining state-of-the-art techniques from both fields, we offer an efficient and automated solution that surpasses traditional methods, providing deeper insights into educational data and enhancing the overall learning experience.

D. PRACTICAL APPLICATIONS AND EDUCATIONAL IMPACT

The automation and efficiency of the method make it applicable to large-scale educational scenarios, such as learning data analysis in college classrooms, automatic collection of educational data, and personalized learning feedback generation. By automating the extraction and analysis of handwritten notes, this method enables educators to monitor student progress continuously and in real time, making it easier to identify trends in learning patterns and detect knowledge blind spots. Such insights allow teachers to make targeted pedagogical adjustments that directly address student needs, thereby improving the quality of teaching and enhancing student outcomes.

Additionally, this method holds significant potential in educational research by facilitating the collection of vast amounts of learning data across multiple institutions and courses. This data can be invaluable for large-scale educational studies, helping to identify common learning difficulties and uncovering patterns that inform curriculum design. Moreover, the scalability of this approach supports its use in adaptive learning platforms, where personalized feedback can be automatically generated based on individual student notes. This integration of automated analysis into personalized learning not only supports students in self-assessment but also fosters data-driven instructional strategies, promoting advancements in educational data mining and evidence-based teaching practices.

E. FUTURE RESEARCH DIRECTIONS

To further enhance the applicability of this methodology, future research could focus on the following areas:

Extension of Diverse Data Sources: Expanding the application of this method to include a wider variety of learning materials, such as class notes, lab reports, and project submissions, can broaden its utility. Adapting the model to work with different types of handwriting and layout formats will make it more versatile across academic fields. Additionally, incorporating multi-language support could increase its effectiveness in diverse educational settings, allowing it to serve a broader spectrum of students and instructors.

Model Optimization and Adaptive Enhancement: Future research could explore adaptive learning algorithms to improve the model's ability to adjust to various handwriting styles, enhancing the accuracy of content matching across unique student writing patterns. Improved prompt engineering for generative AI models, combined with advances in AI-driven handwriting analysis, could further elevate matching accuracy. Enhanced algorithms may allow the model to distinguish subtle differences in content emphasis, enabling more precise feedback.

Real-Time Analysis and Feedback System: Developing a real-time analysis system based on this methodology could allow educators to monitor students' learning status more immediately, enabling timely, data-driven interventions. Such a system would facilitate live analysis during lectures or study sessions, allowing teachers to adjust their instruction dynamically and address student needs as they arise.

Through continued optimization and the integration of these advancements, this method holds promise for making substantial contributions to educational data analytics, adaptive learning, and personalized feedback generation. These improvements would provide educators with smarter, data-driven instructional support, paving the way for more interactive and responsive learning environments.

VI. CONCLUSION

This study proposes an integrated approach for analyzing students' handwritten notes by combining advanced image segmentation and generative AI. By utilizing the Attention Multi-task U-Net model, the method effectively segments various content types, such as text, diagrams, and formulas, and then leverages GPT-4o to extract and analyze key information within each category. The experimental results demonstrate that this method achieves high accuracy in both segmenting and interpreting handwritten notes, providing educators with a robust tool to gain insights into students' learning patterns, knowledge retention, and focus areas. This approach supports a deeper understanding of students' comprehension, helping educators identify concepts that may need further reinforcement.

The primary contributions of this study include the automation of handwritten note analysis, which significantly reduces the time and effort required for educators to manually review and interpret student notes. This is especially valuable in large-scale educational settings where manual review of handwritten notes is often impractical. By incorporating advanced AI techniques, the method achieves a high degree of content matching accuracy, enabling the extraction of meaningful data that can be used for personalized and adaptive learning insights. The ability to accurately link note content to curriculum topics allows educators to provide targeted feedback, helping students strengthen their grasp of core concepts and address any identified knowledge gaps.

Another notable aspect of this methodology is its scalability. The method's high accuracy and adaptability make it suitable for large-scale applications across diverse educational scenarios, from secondary education to university-level courses. By providing actionable data, this approach empowers educators to make data-informed adjustments to their teaching strategies and better support student success. Additionally, the method could be integrated into adaptive learning platforms, offering students a continuous assessment tool that helps guide their studies and aligns with their individual learning paths.

Future Impact: The successful integration of image processing and generative AI for educational data analysis represents a significant advancement in applying AI to adaptive learning environments. As educational technology continues to evolve, this methodology has the potential to expand beyond handwritten notes, analyzing other forms of student-generated data, such as lab reports, project submissions, and digital assignments. This expansion would provide a more comprehensive view of student engagement and learning outcomes, further supporting data-driven instructional methods. Future research could focus on optimizing this method into a real-time feedback system, allowing for immediate insights into student progress and enabling responsive teaching adjustments. Additionally, enhancing the model's adaptability to handle diverse handwriting styles and expanding its scalability for broader educational applications will be essential in maximizing its impact in the field of education.

By continually advancing and refining these AI-driven approaches, this methodology could significantly contribute to the future of personalized education, making it possible for educators to support students in a way that aligns with their unique learning needs and goals.

REFERENCES

[1] X. Li, Y. Zhou, Z. Pan, and J. Feng, "Partial order pruning: For best speed/accuracy trade-off in neural architecture search," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9145–9153.

[2] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau, and C. Wolf, "Residual conv-deconv grid network for semantic segmentation," in *Proc. Brit. Mach. Vis. Conf.*, 2017, pp. 1–12.

[3] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4151–4160.

[4] K. Chen, H. Wei, J. Hennebert, R. Ingold, and M. Liwicki, "Page segmentation for historical handwritten document images using color and texture features," in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit.*, Sep. 2014, pp. 488–493.

[5] J. Jiang, Y.-C. Hu, C.-J. Liu, D. Halpenny, M. D. Hellmann, J. O. Deasy, G. Mageras, and H. Veeraraghavan, "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 134–144, Jan. 2019.

[6] Y. Zhou, T. Minematsu, and A. Shimada, "Improvement of image segmentation model for handwritten notebook analytics," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Kuala Lumpur, Malaysia, Oct. 2023, pp. 1870–1874, doi: 10.1109/ICIP49359.2023.10222873.

[7] E. E. Enekwechi and M. C. Ezeanya, "Study-habit as a predictor of secondary school students' achievement in chemistry in Anambra state," *Int. J. Educ. Eval.*, vol. 7, no. 2, pp. 1–7, 2021.

[8] B. Flanagan, W. Chen, and H. Ogata, "Joint activity on learner performance prediction using the BookRoll dataset," in *Proc. Int. Conf. Comput. Educ.*, 2018, pp. 480–485.

[9] T. Fukawa-Connelly, K. Weber, and J. P. Mejía-Ramos, "Informal content and student note-taking in advanced mathematics classes," *J. Res. Math. Educ.*, vol. 48, no. 5, pp. 567–579, Nov. 2017.

[10] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–12.

[11] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 1856–1867, Jun. 2020.

[12] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1055–1059.

[13] D. He, S. Cohen, B. Price, D. Kifer, and C. Lee Giles, "Multi scale multi-task FCN for semantic page segmentation and table detection," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, 2017, pp. 254–261.

[14] A. Y. Q. Huang, O. H. T. Lu, J. C. H. Huang, C. J. Yin, and S. J. H. Yang, "Predicting students' academic performance by using educational big data and learning analytics: Evaluation of classification methods and learning logs," *Interact. Learn. Environments*, vol. 28, no. 2, pp. 206–230, Feb. 2020.

[15] C. Yang and F. Gao, "EDA-Net: Dense aggregation of deep and shallow information achieves quantitative photoacoustic blood oxygenation imaging deep in human breast," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Jan. 2019, pp. 246–254.

[16] R. S. Jansen, D. Lakens, and W. A. IJsselsteijn, "An integrative review of the cognitive costs and benefits of note-taking," *Educ. Res. Rev.*, vol. 22, pp. 223–233, Nov. 2017.

[17] W. Jia, Z. Zhong, L. Sun, and Q. Huo, "A CNN-based approach to detecting text from images of whiteboards and handwritten notes," in *Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Aug. 2018, pp. 1–6.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[19] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[20] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[21] A. Kundu and T. Bej, "Mobile phone vs paper-pencil mode of note-taking, a case study among the students of Bankura University," *Interact. Technol. Smart Educ.*, vol. 17, no. 1, pp. 1–13, Aug. 2019.

[22] V. Zyuzin and T. Chumarnaya, "Comparison of UNet architectures for segmentation of the left ventricle endocardial border on two-dimensional ultrasound images," in *Proc. Ural Symp. Biomed. Eng., Radioelectron. Inf. Technol. (USBEREIT)*, Apr. 2019, pp. 110–113.

[23] J. Lee, H. Hayashi, W. Ohyama, and S. Uchida, "Page segmentation using a convolutional neural network with trainable co-occurrence features," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 1023–1028.

[24] P. A. Mueller and D. M. Oppenheimer, "The pen is mightier than the keyboard: Advantages of longhand over laptop note taking," *Psychol. Sci.*, vol. 25, no. 6, pp. 1159–1168, 2014.

[25] I. Mwalumbwe and J. S. Mtebe, "Using learning analytics to predict students' performance in Moodle learning management system: A case of Mbeya University of Science and Technology," *Electron. J. Inf. Syst. Developing Countries*, vol. 79, no. 1, pp. 1–13, Mar. 2017.

[26] P. Nilsson, "An inferentialist perspective on how note-taking can constrain the orchestration of math-talk," *Scandin. J. Educ. Res.*, vol. 63, no. 7, pp. 1121–1133, Nov. 2019.

[27] H. Cui, X. Liu, and N. Huang, "Pulmonary vessel segmentation based on orthogonal fused U-Net++ of chest CT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Jan. 2019, pp. 293–300.

[28] H. Ogata, C. Yin, M. Oi, F. Okubo, A. Shimada, K. Kojima, and M. Yamada, "E-Book-based learning analytics in university education," in *Proc. Int. Conf. Comput. Educ. (ICCE)*, 2015, pp. 401–406.

[29] T. Pavlidis and J. Zhou, "Page segmentation and classification," *CVGIP, Graph. Models Image Process.*, vol. 54, no. 6, pp. 484–496, Nov. 1992.

[30] A. Piolat, T. Olive, and R. T. Kellogg, "Cognitive effort during note taking," *Appl. Cognit. Psychol.*, vol. 19, no. 3, pp. 291–312, Apr. 2005.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[32] N. Stamatopoulos, B. Gatos, G. Louloudis, U. Pal, and A. Alaei, "ICDAR 2013 handwriting segmentation contest," in *Proc. 12th Int. Conf. Document Anal. Recognit.*, Aug. 2013, pp. 1402–1406.

[33] A. L. H. van der Meer and F. R. van der Weel, "Only three fingers write, but the whole brain works: A high-density EEG study showing advantages of drawing over typing for learning," *Frontiers Psychol.*, vol. 8, May 2017, Art. no. 706.

[34] W. Wang, E. Xie, X. Li, W. Hou, T. Lu, G. Yu, and S. Shao, "Shape robust text detection with progressive scale expansion network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9336–9345.

[35] B. Li, T. Minematsu, Y. Taniguchi, F. Okubo, and A. Shimada, "How does analysis of handwritten notes provide better insights for learning behavior?" in *Proc. 12th Int. Learn. Anal. Knowl. Conf. (LAK)*, Mar. 2022, pp. 549–555.

[36] K. He, C. Lian, B. Zhang, X. Zhang, X. Cao, D. Nie, Y. Gao, J. Zhang, and D. Shen, "HF-UNet: Learning hierarchically inter-task relevance in multi-task U-Net for accurate prostate segmentation in CT images," *IEEE Trans. Med. Imag.*, vol. 40, no. 8, pp. 2118–2128, Aug. 2021.

[37] R. De Feo, A. Shatillo, A. Sierra, J. M. Valverde, O. Gröhn, F. Giove, and J. Tohka, "Automated joint skull-stripping and segmentation with multi-task U-Net in large mouse brain MRI databases," *NeuroImage*, vol. 229, Apr. 2021, Art. no. 117734.

[38] X. Lu, Y. Zhong, Z. Zheng, Y. Liu, J. Zhao, A. Ma, and J. Yang, "Multi-scale and multi-task deep learning framework for automatic road extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9362–9377, Nov. 2019.

[39] X. Yang, Z. Zeng, S. Y. Yeo, C. Tan, H. L. Tey, and Y. Su, "A novel multi-task deep learning model for skin lesion segmentation and classification," 2017, *arXiv:1703.01025*.

[40] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019.

[41] J. Dolz, I. B. Ayed, and C. Desrosiers, "Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities," in *Proc. Int. MICCAI Brainlesion Workshop*, vol. 11383, Jan. 2018, pp. 271–282, doi: 10.1007/978-3-030-11723-8_27.

[42] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie, "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Med. Phys.*, vol. 46, no. 2, pp. 576–589, Feb. 2019, doi: 10.1002/mp.13300.

[43] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Cham, Switzerland: Springer, Jul. 2017, pp. 506–517, doi: 10.1007/978-3-319-60964-5_44.

[44] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "MDU-Net: Multiscale densely connected U-Net for biomedical image segmentation," *Health Inf. Sci. Syst.*, vol. 11, Sep. 2023, Art. no. 13.

[45] G. Du, X. Cao, J. Liang, X. Chen, and Y. Zhan, "Medical image segmentation based on U-Net: A review," *J. Imag. Sci. Technol.*, vol. 64, no. 2, Mar. 2020, Art. no. 020508, doi: 10.2352/j.imagingsci.technol.2020.64.2.020508.

[46] U. G. Naidu, R. Thiruvengatanadhan, S. Narayana, and P. Dhanalakshmi, "Character level segmentation and recognition using CNN followed random forest classifier for NPR system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, pp. 12–18, 2022, doi: 10.14569/ijacsa.2022.0131102.

[47] J. Hong, B.-Y. Park, M. J. Lee, C.-S. Chung, J. Cha, and H. Park, "Two-step deep neural network for segmentation of deep white matter hyperintensities in migraineurs," *Comput. Methods Programs Biomed.*, vol. 183, Jan. 2020, Art. no. 105065.

[48] J. Jo et al., "Handwritten text segmentation via end-to-end learning of convolutional neural networks," *Multimedia Tools Appl.*, vol. 79, pp. 32137–32150, 2020. [Online]. Available: https://doi.org/10.1007/s11042-020-09624-9

YUNYU ZHOU received the M.E. degree in engineering from Chongqing University, in 2017. He is currently pursuing the Ph.D. degree in engineering with Kyushu University.

His Ph.D. research focuses on advancing methodologies in image processing and integrating them with artificial intelligence systems to address complex challenges in education and beyond. He is particularly passionate about leveraging generative AI models to enhance the interpretation of visual data and to develop novel applications in personalized education. Through his research in learning analytics, he aims to improve the understanding of student behavior and optimize learning outcomes using data-driven approaches. His research interests include image processing, learning analytics, and generative AI.

CHENG TANG (Member, IEEE) received the Ph.D. degree in engineering from the University of Toyama, Toyama, Japan, in 2022. From 2022 to 2023, he was an Assistant Professor with the Graduate School of Engineering, Tsukuri College, Nagoya Institute of Technology, Nagoya, Japan. In 2023, he joined Kyushu University, Fukuoka, Japan, where he is currently an Assistant Professor (Tenured) with the Department of Advanced Information Technology, Graduate School and Faculty of Information Science and Electrical Engineering. His main research interests include medical data mining, time series forecasting, computer vision, natural language processing, and learning analytics.

ATSUSHI SHIMADA (Member, IEEE) received the M.E. and D.E. degrees from Kyushu University, in 2007. He is currently a Professor with the Faculty of Information Science and Electrical Engineering, Kyushu University. He was a JST-PRESTO Researcher, from 2015 to 2019. His current research interests include learning analytics, pattern recognition, media processing, and image processing. He received the MIRU Interactive Presentation Award, in 2011 and 2017, the MIRU Demonstration Award, in 2015, the Background Models Challenge 2012 The First Place, in 2012, the PRMU Research Award, in 2013, the SBM-RGBD Challenge The First Place, in 2017, the ITS Symposium Best Poster Award, in 2018, the JST PRESTO Interest Poster Award, in 2019, the IPSJ/IEEE-Computer Society Young Computer Researcher Award, in 2019, the CELDA Best Paper Award, in 2019, and the MEXT Young Scientist Award, in 2020.