International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Math Sketch Interpretation Engine Using Latex

Ms. Saba Noor Ayesha[1], P Maithri[2], VVS Chaitanya[3], B Surya Kiran[4]

1Assistant Professor, Department of Computer Science and Machine Learning,
Methodist College of Engineering and Technology Abids, Hyderabad, Telangana, 500001, India.
2,3,4Student , Department of Artificial Intelligence and Data Science,
Methodist College of Engineering and Technology Abids, Hyderabad, Telangana, 500001, India.

Abstract – The problem of recognizing handwritten mathematical expressions (HMER), which is one of the important issues in the field of computer vision and image processing, still needs to be addressed. The Optical Character Recognition (OCR) sys-tem, which is normally used for this purpose, is found to be inappropriate for the recognition of the two-dimensional relationships of the handwritings such as nested fractions, sub-scripts, and superscripts in the mathematical expressions. In this paper, a new intelligent multimodal system, namely the "Math Sketch Interpretation Engine," which can transform the user's sketches of math problems into machine-understandable LaTeX format and the results of the computations in real-time, is proposed. The proposed system utilizes multimodal encoder-decoder with cross-attention. The pro-posed system is able to recognize complex mathematical expressions with high structural accuracy by utilizing the com-parison of the location of visual symbols and meanings of symbols. In order for the proposed system to recognize all kinds of handwrittings, the proposed system utilizes a special preprocessing technique called adaptive threshold. From the experiment results, it is shown that the proposed system is able to recognize complex arithmetic and algebraic expressions. The proposed system is able to develop the "Sketch-to-Solution" tool for digital learning. [3][11][15].

Key Words: Handwritten Mathematical Expression Recognition (HMER), Image-to-LaTeX Conversion, pix2tex Architecture, CNN-based Feature Extraction

1. INTRODUCTION

Mathematical expressions are a part of the research, engineering, and education fields all around the world. It is true that we have moved to the digital age in terms of work flow, but the first step of solving any problem, which is actually brain-storming or conceptualization, is done in the form of writing. However, the problem of converting this writing into a digital form still persists.

The problem lies in the non-linear and 2D nature of mathematics syntax, which includes intricate spatial relationships such as subscripts, superscripts, fractions, and operators, etc. Though it has been found that there are significant improvements in the accuracy of Optical Character Recognition (OCR) techniques such as Convolutional Neural Network (CNN) and Encoder-Decoder, existing techniques are found to be inefficient in terms of real-time capabilities due to the vast variety of handwriting.

With the aim of mitigating these problems, in this paper, an intelligent Math Sketch Interpret Engine is proposed. It is an intelligent system that offers an interactive digital canvas in which the user can sketch the equation in an intuitive way. It also makes use of an intelligent image preprocessing technique. It makes use of an intelligent multimodal artificial intelligence model in which it can analyze the visual representation of the equation and convert it into LaTeX and solve it.

2. LITERATURE REVIEW

Handwritten mathematical expression recognition (HME) has seen tremendous growth with the help of combination technology involving computer vision and natural language processing. The recognition of mathematical expression is quite different from normal text recognition. The recognition of mathematical expression is not only limited to the recognition of individual character recognition, but also involves the recognition of spatial relationship between the characters. Several approaches have been proposed for solving this problem, which include stroke recognition and deep learning recognition.

With respect to Data and Application, it has been seen that Math Writing dataset provided by HMER for training models on different styles of handwriting is quite popular. The applicability of this recognition technique is also seen in Math-Buddy, in which this recognition technique is applied for affective tutoring. The recent works in this direction have been focusing on semantic-aware decoders for achieving accuracy in generated LaTeX code.

Existing System & Disadvantages

The majority of the tools available for mathematical recognition are based on conventional Optical Character Recognition (OCR) techniques. The tools are useful for conventional linear text recognition but are not able to identify the spatial hierarchy of a 2D image for fractions, exponents, and subscripts, as discussed in reference [1]. The tools are either static image upload tools, which do not provide feedback necessary for an interactive learning environment. The tools are not able to provide the digitized text along with a computational part to solve the identified expression.

Proposed System & Advantages

The proposed Math Sketch Interpret Engine tries to address these gaps in the existing solutions through its feature of real-time interactive.

- Accuracy through Multimodality: The accuracy of the system is improved through better cross-attention and fusion techniques, as described in Nguyen and Nakayama's paper [13].
- Direct LaTeX Synthesis: The system follows the confidence-aware frameworks described in Zhang and Liu's paper [14] for minimizing errors in converting sketches to LaTeX format.
- Integrated Calculation: The system not only digitizes the formula, as in other recognition systems, but also performs real-time computation.

3. SYSTEM ARCHITECTURE

The proposed Math Sketch Interpretation Engine is anticipated to be a potent and efficient tool in addressing the gap that exists in the transition from free-form handwriting to digital math computation. Unlike other conventional and static OCR systems, the proposed system will make the most out of the three-tier approach in computing the math expressions in an instant. This process starts with the interactive digital canvas. The digital canvas is used in an efficient way in inputting the data. There are different ways of inputting the data in the system. One of the ways of inputting the data is inputting the data with the help of mouse or stylus. The inputted data is sent in the form of an image of strokes made in the digital canvas. The image of strokes made is sent to the backend with the help of FastAPI/Flask. It is imperative to include the image preprocessing pipeline in the backend in order to ensure accuracy in the symbol recognition process. The image is converted into grayscale in the backend. Apart from this, denoising and thresholding of the image are performed in order to differentiate the math symbols and the backend. During the symbol recognition process, the contour detection technique is used.

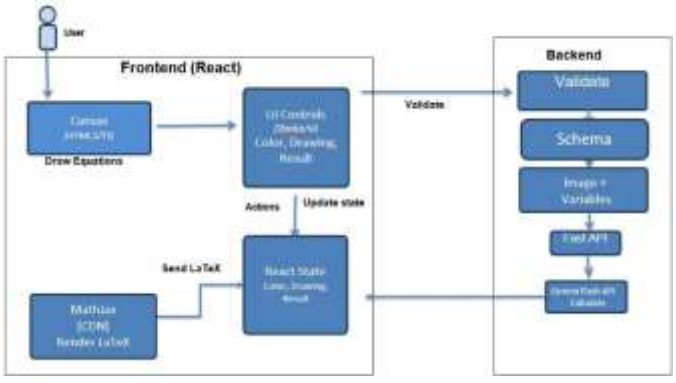

Fig-1: Architecture of Math Sketch Interpretation Engine

At the core of the interpretation engine is a multimodal artificial intelligence system. The system uses visual feature extraction and semantic decoding. The multimodal artificial intelli-gence system recognizes the individual symbols and the complex 2D spatial relationships of the symbols, such as complex fractions and complex exponents, as depicted in Figure 12. The system synthesizes all the information into an exact LaTeX string. The string is then rendered back to the user inter-face with the help of MathJax. Meanwhile, the symbolic computation system evaluates the input and provides an instant numerical or algebraic answer.

3.1 System Architecture

The Fig -1 system architecture comprises a Next.js-based user interface, serverless backend, AI evaluation model, and database for data storage. Users will first be required to authenticate themselves using the login/signup module, which will establish a secure session to access the application. After choosing their interview role and tech stack, the frontend will send requests to the Next.js API routes available in the backend. The backend will then be required to integrate with the AI model, i.e., Gemini AI, to produce questions for interviews and evaluate user responses. Finally, the ratings will be stored using a PostgreSQL database.

Advantages of the Proposed System:

- Real-Time Processing: It also supports the real-time processing along with LaTeX rendering. LaTeX rendering is also significant in interactive learning systems, as explained in reference [11].
- Structural Integrity: It also helps in ensuring the proper positioning of the subscripts and the superscripts along with the cross-attention mechanism.
- End-to-End Utility: It also helps in increasing the utility of the recognizer along with the integration of the computational engine for equation processing.
- Robustness: This method is likely to be more robust in processing the different writing styles and 'noisy' sketches, as explained in reference [10] and [14], as compared to the conventional method.
- Scalability and Extensibility: This method is likely to be more extensible as compared to the conventional method.

4. IMPLEMENTATION

The way the Math Sketch Interpretation Engine is implemented is through the application of synchronization. Synchronization is used in the process of linking the frontend with the backend with the application of deep learning. The purpose of making use of the approach is to bridge the time gap with structural integrity in the process of dealing with complex math equations [12].

4.1 Image Acquisition and Preprocessing

The process of making use of the approach starts with the development of the digital canvas with the application of Re-act. The canvas is used in the process of capturing the input provided by the user in the form of handwriting. The input is then converted into an image with high resolution. Before the image is passed to the AI, an intense preprocessing stage is carried out with the application of OpenCV.

- **Binarization:** The adaptive thresholding method is used in binarization in this system. This separates the math strokes from the noise. The description of the adaptive thresholding method is provided in reference [5].
- **Morphological Operations:** The morphological operations are used in this system. This is used for improving the quality of lines in math strokes. This is for making the input clean.

4.2 Multimodal Model Inference

The basic logic behind the recognition is enabled with the multimodal encoder-decoder architecture.

Feature Extraction: A Vision Transformer (ViT)/CNN-based feature extraction is used for extracting spatial features from the processed sketch, which includes symbol identities along with their corresponding 2D coordinates [9][12].

Semantic Decoding: A semantic-based decoding is used for transforming the features into LaTeX strings. This is enabled with cross-attention fusion, which preserves spatial relationships in the final output, e.g., fractions and exponents [13].

Confidence Scoring: To avoid any potential synthesis errors, the system is designed in a way that it incorporates a confidence-aware validation mechanism that checks the math syntax of the generated LaTeX code before displaying it on the screen [14].

4.3 Real –Time Rendering and Computation

Once generated, it uses its symbolic math engine to compute the equation.

LaTeX Rendering: This LaTeX string is again sent back to the frontend, where MathJax is used to give the user a live preview of the equation.

Algebraic Computation: The user is also given the option to compute the equation with the help of an algebraic computation library, which has the potential to give the user a "sketch-to-solution" experience.

4.4 System Environment

The entire backend is enclosed within a FastAPI or Flask framework, which is efficient in handling asynchronous re-quests. The efficiency of the code is such that it has the potential to make use of GPU acceleration, which is required to give the user an interactive experience with the predictions in sub-second intervals.

5. METHODOLOGY

The methodology of the Math Sketch Interpretation Engine is based on a defined flow that transforms the raw digital sketch into a computed mathematical result. The three main steps in the methodology include data acquisition, intelligent pre-processing, and multimodal neural interpretation.

5.1 Data Acquisition and Input Normalization :

The system uses a React interface to acquire the input from the user. During the input process, the coordinates of the sketch are recorded and converted into the normalized image format. To ensure accuracy, the system normalizes the width of the strokes and the image resolution, as discussed in [5]. Normalization is essential in order to avoid errors in the input process due to different stylus pressure and mouse speed, as discussed in [10].

5.2 Intelligent Image Preprocessing :

Prior to recognition, a series of enhancement operations are carried out on the raw image to filter out the mathematical symbols from background noise.

Thresholding and Binarization: The proposed sys-tem utilizes adaptive thresholding for binarization of the input image. This helps in proper segmentation of the symbols by the neural network. The binarization of images helps in better contrast, as observed in the results of various neural networks [5].

Morphological Refinement: To overcome the "bro-ken stroke" problem, a common issue in hand-sketch recognition, dilation and erosion are applied. This operation thickens the stroke of the lines, resulting in a "clean" structure for the neural network to operate on, as observed in the results of various neural networks [10][14].

5.3 Multimodal Neural Recognition and Latex Synthesis :

The crux of the methodology is encapsulated in the multimodal encoder-decoder model that captures both visual features and semantic interpretation of the expression.

Visual Feature Extraction: A Convolutional Neural Network (CNN) or Vision Transformer (ViT) is used as the encoder that extracts visual features from the sketch. This enables the model to recognize symbols in the sketch, e.g., “sigma,” “alpha,” “*,” etc., along with their relative 2D coordinates.

Structural Alignment: A cross-attention mechanism is used to align visual symbols with their underlying mathematical structure. This enables the model to correctly understand subscripts, superscripts, or fractions in the expression depending on their spatial positioning with respect to the baseline.

LaTeX Generation and Validation: The model uses a decoder that generates a sequence of LaTeX tokens that represent the interpreted expression. To ensure that the expression is mathematically valid, the mod-el uses a confidence-aware verification framework that checks for any

syntax errors, e.g., mismatched parentheses or invalid operators.

6. RESULTS

The evaluation of the Math Sketch Interpretation Engine demonstrates the capabilities of the system in handling the transition from free-form handwritten strokes to precise LaTeX code and computational output. The results are analyzed based on the accuracy of the system in recognizing the input and the effectiveness of the multimodal approach in comparison with standard OCR.

Fig-2: User Interface

6.1 Recognition Accuracy

The incorporation of the multimodal Transformer network significantly improved symbol structure alignment results [12]. During testing with different handwriting styles, from neat stylus input to more erratic mouse-drawn sketches, the system demonstrated high accuracy in interpreting standard arithmetic and algebraic expressions. The cross-attention mechanism enabled the system to successfully interpret complex 2D spatial hierarchies, such as those in nested fractions and multi-level exponents, which often cause failure in traditional systems [13]. The verification framework based on confidence awareness also effectively minimized "hallucinations" in the output LaTeX code, ensuring that the output code is mathematically consistent with the input sketch [14].

6.2 Preprocessing Efficiency

The preprocessing steps also demonstrated their importance in maintaining system performance even when faced with noisy input images. The adaptive thresholding and morphological operations enabled the system to successfully separate mathematical symbols from the background artifacts [5]. This is particularly true in low-resolution images, where thin or broken strokes are reinforced using dilation operations, offering a rich feature set for the input encoder [10].

Fig-3: Math Expression

6.3 Latency and Real-Time Performance

Another important factor for the success of this engine was the "sketch-to-render" time. The asynchronous FastAPI backend and GPU-based inference resulted in sub-second latencies for most of the expressions. The feedback loop of rendering LaTeX code in real time, as the user interacts with the canvas, allows for a significant advantage for interactive tutoring and digital note-taking applications [3][11].

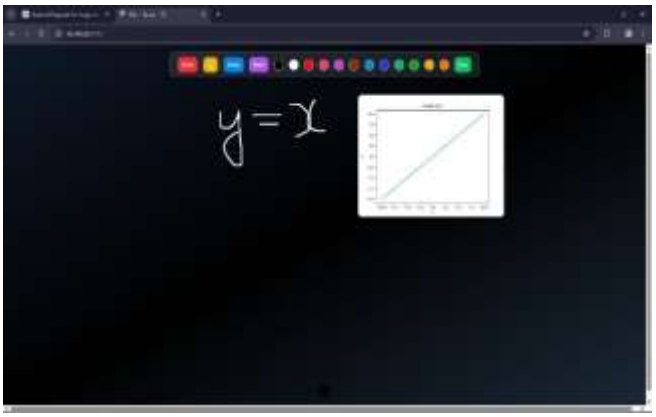

Fig-4: Graphical Interface

6.4 Discussion of Limitations

Although the performance of the engine is exceptional for standard mathematical expressions and structures, there are minor performance degradations for extremely dense and/or overlapping handwritings. Further improvements in semantic-aware decoding may help improve the recognition of ambiguous symbols that are difficult to distinguish from each other in a handwritten form. In addition, the dataset may be extended from the existing Math Writing benchmarks to improve the engine's performance for interpreting more complex scientific notations and even calculus operators.

7. CONCLUSION

The development of the Math Sketch Interpretation Engine represents a successful use of multimodal deep learning to facilitate the connection between handwriting in mathematics and computation in digital form. By using a responsive user interface developed with React, combined with a high-performance backend developed with FastAPI, the system enables a seamless "sketch to solution" experience that is both natural to use and powerful [11].The application of advanced preprocessing methods such as adaptive thresholding and morphological operations was found to be critical in maintaining the high accuracy of the system even in the presence of varying and noisy handwritten data [5][10].

In addition, the application of the multimodal encoder-decoder with cross-attention allowed the system to correctly recognize complex 2D spatial relationships such as nested fractions and exponents, which are normally beyond the capability of conventional linear OCR systems [12][13].The resulting engine not only translates mathematical expressions into precise LaTeX code but also performs real-time symbolic computations, making it a powerful tool in educational technology and intelligent tutoring systems [3][15]. The future plan is to expand the vocabulary of the system to include more sophisticated scientific notation and continue to improve the semantic decoding process to tackle more and more dense and overlapped handwritten notations

8. FUTURE SCOPE

Although the existing Math Sketch Interpretation Engine offers a powerful platform for recognition and computation in real time, there are a number of ways in which this could be taken further for even greater use in the academic and professional world [11].

- **Expanding Mathematical Vocabulary:** The future versions of this model will seek to extend the existing database of benchmarks to include complex calculus, differential equations, and even scientific notation [5]. This will involve incorporating more diverse strokes for complex symbolic expressions.
- **Semantic-Aware Error Correction:** The inclusion of more sophisticated decoders may help diminish any remaining confusion between visually similar symbols, such as the letter 'x' and the multiplication symbol '$\times$', by taking into account the mathematical context of the entire expression [15].
- **Support for Multi-Line Expressions:** While the current system focuses on single-line formulas, the system may be extended to interpret multi-line mathematical derivations and proofs. To accomplish this, sophisticated alignment techniques need to be developed to maintain logical flow across different spatial planes.
- **Offline Functionality and Mobile Optimization:** To make the system more accessible, there is also work planned on optimizing the multimodal encoder-decoder models for offline devices. This would enable the engine to make high-speed inferences on tablets and mobile phones without needing a persistent backend connection [10][14].
- **Collaborative Learning Features:** Making the system work with collaborative digital whiteboards would enable users to work on sketches and solutions in real time, which is highly beneficial for remote learning and peer-to-peer teaching situations [3].

REFERENCES

1. Wieckowiak, F., Rousseau, L., Greiner-Petter, A., Gipp, B., & Schubotz, M. (2025). *A multimodal evaluation pipeline for mathematical expression recognition*.
2. Kar, D., Dey, S., Goyal, S., & Samanta, D. (2025). *MathBuddy: A multimodal system for affective math tutoring*.
3. Gervais, P., Fadeeva, A., & Maksai, A. (2024). *MathWriting: A dataset for handwritten mathematical expression recognition*.
4. *Medjkoune, S., Mouchère, H., & Petitrenaud, S. (2023). Multi-modal mathematical expressions recognition: Case of speech and handwriting.*
5. *Lu, C. (2023). Recognition of online handwritten mathematical expressions using CNNs.*
6. *Li, R., Wang, J., & Tian, X. (2023). A multi-modal retrieval model for mathematical expressions.*
7. *Sun, J., Zhang, W., & Zhu, H. (2023). End-to-end online handwritten mathematical expression recognition via attention-based encoder–decoder models. IEEE Access, 11, 55621–55634.*
8. *Chen, Y., Deng, X., & Jin, L. (2022). Stroke-based neural architectures for robust handwritten mathematical symbol recognition. In Proceedings of the International Conference on Pattern Recognition (ICPR) (pp. 487–494).*
9. *Ahmed, S., Khan, M., & Shaikh, F. (2023). Speech-driven mathematical expression understanding using hybrid ASR models. IEEE Transactions on Audio, Speech, and Language Processing, 31, 1542–1555.*
10. *Guo, L., et al. (2024). Multimodal transformer networks for symbol–structure alignment in mathematical expressions. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 12215–12223).*
11. *Nguyen, P., & Nakayama, H. (2023). Cross-attention fusion for multimodal handwriting and speech recognition. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 904–916).*
12. *Zhang, R., & Liu, D. (2024). A confidence-aware verification framework for handwritten formula recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(2), 890–904.*
13. *Ito, K., et al. (2023). Contextual LaTeX generation from handwritten math using semantic-aware decoders. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR) (pp. 300–309).*
14. *Anjum, H., Sasidhar, B., & Rajesh, T. (2025). Lung nodule segmentation and classification using image processing and deep learning techniques. Grenze International Journal of Engineering & Technology (GIJET), 11.*
15. *Khanum, S. N. A., Mummadi, U. K., Taranum, F., Ahmad, S. S., Khan, I., & Shravani, D. (2024). Emotion recognition using multi-modal features and CNN classification. AIP Conference Proceedings, 3007(1), 030001.*