IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 12, 2019, accepted September 24, 2019, date of publication October 7, 2019, date of current version October 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2945825

Detecting Mathematical Expressions in Scientific Document Images Using a U-Net Trained on a Diverse Dataset

WATARU OHYAMA[1], (Member, IEEE), MASAKAZU SUZUKI[2], AND SEIICHI UCHIDA[3], (Member, IEEE)

[1]Graduate School of Engineering, Saitama Institute of Technology, Fukaya-shi 3690293, Japan
[2]Faculty of Mathematics, Kyushu University, Fukuoka 8190395, Japan
[3]Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 8190395, Japan

Corresponding author: Wataru Ohyama (ohyama@sit.ac.jp)

This work was supported in part by JSPS KAKENHI under Grant JP17H06100.

ABSTRACT A detection method for mathematical expressions in scientific document images is proposed. Inspired by the promising performance of U-Net, a convolutional network architecture originally proposed for the semantic segmentation of biomedical images, the proposed method uses image conversion by a U-Net framework. The proposed method does not use any information from mathematical and linguistic grammar so that it can be a supplemental bypass in the conventional mathematical optical character recognition (OCR) process pipeline. The evaluation experiments confirmed that (1) the performance of mathematical symbol and expression detection by the proposed method is superior to that of InftyReader, which is state-of-the-art software for mathematical OCR; (2) the coverage of the training dataset to the variation of document style is important; and (3) retraining with small additional training samples will be effective to improve the performance. An additional contribution is the release of a dataset for benchmarking the OCR for scientific documents.

INDEX TERMS Character recognition, neural networks, object detection.

I. INTRODUCTION

The performance and effectiveness of document retrieval systems heavily depend on both the amount and quality of registered document content. Although born-digital documents have become more common recently, a large number of printed documents remain. To input such printed documents into retrieval systems, optical character recognition (OCR) techniques have been used for digitizing documents for a long time. Continuous research and development over the last five decades have achieved OCR techniques that are sufficiently mature for such a purpose.

Although OCR techniques demonstrate good performance for digitizing ordinary text in documents, there is still scope for improvement in the recognition accuracy of mathematical expressions (MEs). MEs are essential information containers, particularly for scientific articles and textbooks. The accurate recognition of MEs is strongly expected because it has a wide variety of applications, for instance, correct retrieval, automatic proofing of MEs, and learning support for blind or handicapped people. ME recognition has been considered and developed as an independent module outside of ordinary OCR because of the distinctive properties of MEs, where spatial structures and spatial relationships between symbols contain mathematical information.

Zanibbi and Blostein [1] stated that there are four key problems in ME recognition: ME detection, symbol extraction and recognition, layout analysis and mathematical content interpretation. These four problems are closely related to each other. In particular, ME detection has a large influence on other tasks. There are two types of MEs: displayed and in-line, as shown in Figure 1. The detection processes for each displayed (offset from text lines) and in-line (embedded in text lines) ME are usually implemented separately. When an in-line ME is not detected, the expression is passed to the recognition module for ordinary characters even though it should be passed the recognition module for mathematical symbols. This scenario commonly occurs, and the undetected ME may cause recognition errors that cannot be easily corrected in subsequent postprocessing modules.

Because of the high performance of deep neural networks (DNNs), particularly convolutional neural

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

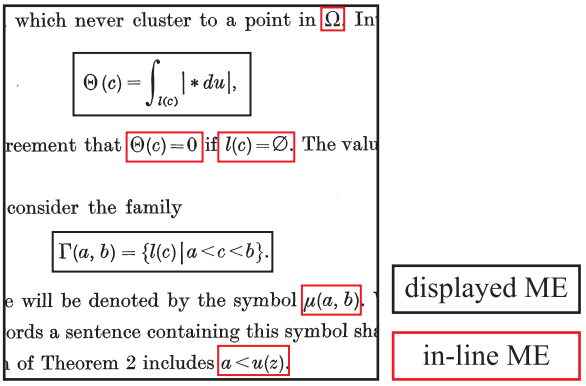


FIGURE 1. Examples of displayed MEs (marked by the black boxes) and in-line MEs (marked by the red boxes). The displayed MEs have offset spaces from the text lines. The in-line MEs are embedded in the text lines.

networks (CNNs) for several image recognition tasks, the DNN architecture is expected to demonstrate promising performance for the problems in document image analysis (DIA), in addition to ME recognition. Some attempts to introduce DNNs for ME recognition have been proposed [2]–[5]. Most of these attempts involved handwritten ME recognition; there is less research on introducing DNNs for ME recognition in printed scientific document images. One example was proposed by He *et al.* [6], who used an end-to-end framework for ME recognition and reported the benchmarking results of the framework. This method can recognize mathematical symbols and the structure of expressions from camera-captured images, which means that it implicitly assumes that the MEs were correctly extracted.

In this paper, we propose a method for extracting MEs from page images of scientific documents using a CNN-based image conversion technique. The proposed method uses a similar architecture to U-Net [7], which was originally developed for biomedical image segmentation. The proposed method solves ME extraction as an image conversion problem, where an original document image is converted to an image that contains only mathematical symbols. Unlike conventional mathematical OCR [8] or ME extraction methods [9], [10], the proposed method does not use any OCR results or a priori knowledge of mathematics and languages.

Figure 2 shows the standard process pipeline for the recognition of scientific documents. As described above, there are two types of MEs: displayed and in-line. The detection process for each of them are implemented separately in the layout analysis, and the line and word segmentation modules, respectively. Therefore, these detection processes typically use different algorithms for each target. The proposed method detects both types of MEs using a unified image transform operation. Consequently, the proposed method can bypass the layout analysis, and line and word segmentation processes so that subsequent ME recognition is simplified by the resultant converted images.

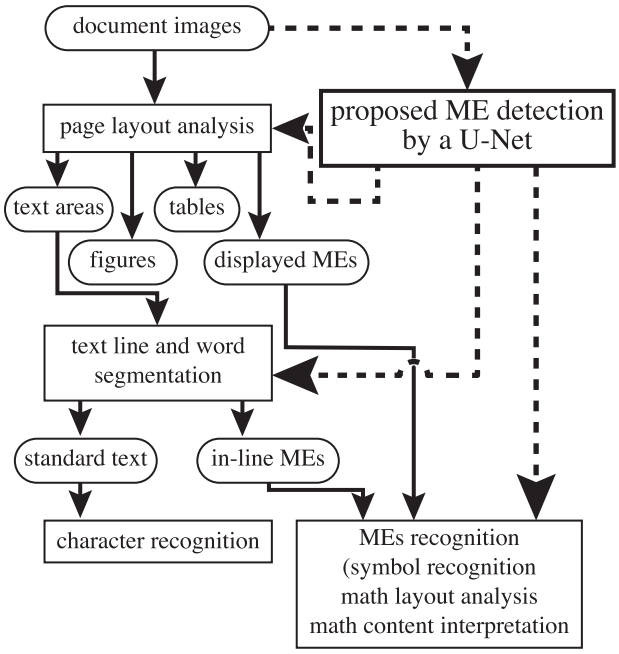


FIGURE 2. The relationship between the proposed method and the standard scientific document recognition pipeline. The standard scientific document recognition pipeline is represented by the solid line arrows. The proposed ME detection can bypass the ME detection process, represented by the broken line arrows, in the page layout analysis, and text line and word segmentation stages. Note that the pre- and postprocesses are omitted in the figure.

The main technical contributions of this research are as follows:

1) We demonstrate that U-Net can achieve very accurate ME extraction when it is trained with a large diversity of training images. To the best of our knowledge, this is the first attempt to apply image conversion using a fully CNN (FCN), including U-Net, to detect MEs in scientific document images.
2) The performance of the proposed method is evaluated using a large and diverse dataset collected from real scientific articles. Analysis from multiple perspective for the behavior of the method is also provided.
3) The proposed method is superior regarding ME detection performance and ease of retraining to InftyReader, which is state-of-the-art software for mathematical OCR.

An additional contribution is that the large annotation dataset that we used for performance analysis has been released for benchmarking OCR performance for scientific documents. Although we could not include the original document images of articles for copyright reasons, we provided hyperlinks to the web pages of the original documents, where readers can obtain the document images.

The remainder of the paper is organized as follows: In Section II, an overview of related works for ME detection is provided. In Section III, the proposed ME detection method using U-Net is described. In Section IV, the performance evaluation is presented. This section includes the dataset,

criteria for performance evaluation and experimental settings. In Sections V and VI, the experimental results and discussion are provided. Finally, in Section VII, the conclusion of the paper is provided.

II. RELATED WORK

A. MATHEMATICAL OCR

Although recent research in ME recognition has concerned the recognition of on-line and off-line handwritten MEs, which is thought to be more difficult than recognizing typeset MEs, there are still problems to solve in the field of mathematical OCR for printed document images. Mathematical OCR poses the following typical difficulties: a large number of character and symbol categories, abnormal (touching and broken) characters, size variation, and complexity of expressions. In the literature, many technologies have been separately proposed for each problem. Garain and Chaudhuri [11] provided a survey of such mathematical OCR technologies for printed document images.

As mentioned in the previous section, mathematical OCR consists of four key problems: ME detection, symbol extraction and recognition, layout analysis and mathematical content interpretation. Among these problems, ME detection is crucial for subsequent process stages. Integrated mathematical OCR systems [8], [12], or attempts to integrate mathematical OCR into conventional OCR systems [10], [13], have been proposed; however, their performance still needs improvement for ME detection. Most recently, there have been application software and libraries, Mathpix [14], for example, that can manage the tasks from symbol recognition to content interpretation. These types of software require that users manually extract MEs with tight bounding boxes.

B. TRADITIONAL METHODS FOR ME DETECTION

As shown in Figure 2, displayed and in-line MEs are extracted at different stages in the pipeline. Most conventional research has proposed extraction methods for either of them or two separate algorithms for each type of ME.

Because displayed MEs are thought to be easier to detect than in-line MEs, rule-based methods were adopted in early research. The earliest work on ME detection, by Lee and Wang [15], classified all text lines into ordinary text or expressions using Bayes decision rules over simple line spacing information. Kacem *et al.* [16] also managed the extraction of displayed MEs using a rule-based algorithm over connected components (CCs) using simple visual features. Das *et al.* [17] divided displayed MEs into multiple categories and built several rules-based decision trees to extract each of them. Chowdhury *et al.* [17] proposed a fully automatic segmentation method for displayed MEs using only the spatial layout information of MEs. Chang *et al.* [18] handled MEs identification as a sub problem of document layout analysis. Other approaches in which trainable classifiers are adopted have been reported. Drake and Baird [19] used neighbor graph representation for CCs in the document image. Phong et al. [20], [21] used a support vector machine over the features in the frequency domain to classify the input text line.

Methods for in-line ME detection are roughly divided into two groups: classification of CCs and postprocessing over the symbol recognition results. The methods in the former group receive CCs extracted using line segmentation and classify the CCs using image features and trained classifiers. Kacem *et al.* [16] proposed a method based on fuzzy logic and a propagation algorithm over adjacent CCs. The method in the postprocessing group applied language and relation models on the results of symbol recognition. Garain [9], [22] applied the symbol *n*-gram model obtained by training data that consisted of a wide corpus for ME recognition. Lee and Wang [15] also managed the in-line ME detection problem as a postprocess following character recognition. In the method, the expression formation process uses a symbol relation tree and decomposition of each text line into a primitive token collection. The method also offers error correction for character recognition results. However, because the method is based on heuristic rules, it is difficult to adopt for a wide variety of scientific documents.

C. DNN-BASED METHODS FOR MATHEMATICAL OCR

Chan [23] and Zanibbi and Blostein [1] described in their survey papers that the recognition systems for MEs are mainly divided into three groups according to their input formats; born-digital vector graphics (PDFs), strokes (handwritings), or document images (static mathematical OCR). Inspired by its significant performance in other image recognition tasks [24]–[29], several attempts which utilize DNN for Mathematical OCR have been reported. However, most of these DNN-based methods handle only the recognition task by assuming that mathematical expressions are extracted by a tight bounding-box using some extraction methods

For born-digital PDF documents, while a method using a combination of hand-crafted features and conventional machine learning algorithm for MEs identification has been proposed by Lin *et al.* [30], some DNN-based methods has been proposed recently for MEs recognition and page segmentation. For instance, Gao *et al.* [31] proposed a deep learning-based MEs detection method for PDF documents. However, since their method relies on the character information embedded in the input PDF file, it is not applicable for static document page images.

For recognizing handwritten MEs (HMEs), recognition methods which handle HMEs as static images [2], [4] and those which utilize online stroke information [3], [32] were proposed. Recently, Zhang *et al.* [5] proposed an end-to-end approach for online HMEs recognition and demonstrated the strong complementarity between offline information with static-image input and online information with ink-trajectory input.

D. U-NET FOR DOCUMENT IMAGE ANALYSIS

The promising performance of FCNs including U-Net for image conversion has encouraged researchers to introduce

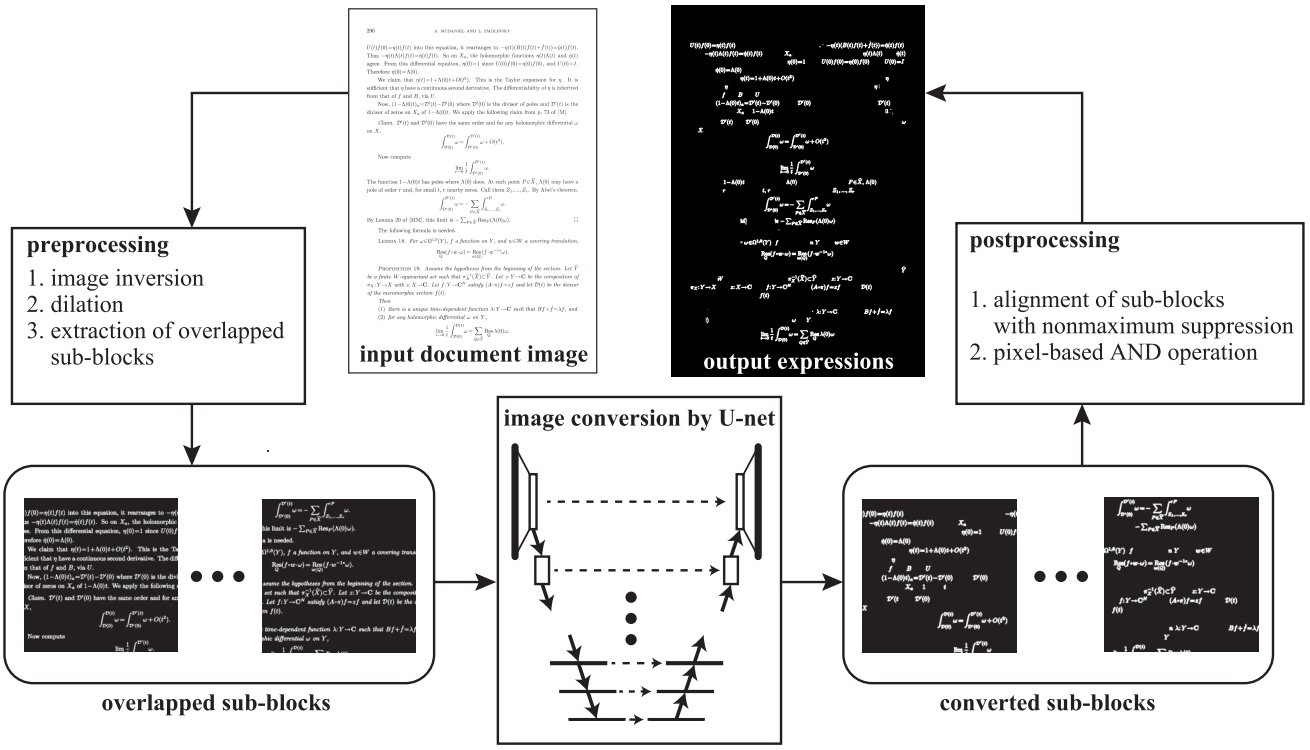


FIGURE 3. Outline of the proposed method. The method consists of three main stages: preprocessing, image conversion by U-Net and postprocessing.

FCN architecture for DIA tasks. Ma *et al.* [33] proposed a method that adopts a stacked architecture of U-Nets to correct the warping distortion of a document image. The binarization of document images is a basic and very important problem in DIA. The report on the document image binarization (DIB) competition [34] remarked that some competitors used U-Net for DIB and obtained high performance. Base line detection [35], text line segmentation [36] and page segmentation [37], [38], which are also common problems in DIA, can be managed using image conversion by U-Net.

These methods also intend to take advantage of U-Net so that it is easy to apply end-to-end training. If a large number of training image datasets that consist of input and desired output images is available, then U-Net is expected to achieve required image conversion from the input to the output.

The proposed method also uses these properties of U-Net for ME detection for printed documents. Additionally, it has the following distinguishing properties from the aforementioned conventional methods: First, the proposed method is based on image conversion from an original document image to an image containing only mathematical symbols. Instead of handcrafted rules for determining MEs, the proposed method uses end-to-end training on a large-scale dataset. Second, the proposed method does not require any mathematical and linguistic knowledge. Third, the proposed method can be embedded in the standard pipeline of ME recognition because it is implemented with no assistance from layout analysis and symbol recognition.

III. MATHEMATICAL EXPRESSION DETECTION BY U-NET

The outline of the proposed method is shown in Figure 3. The proposed method takes a binary document page image as input and outputs an image containing CCs that construct displayed and in-line MEs. The proposed method mainly consists of three stages: (1) preprocessing; (2) image conversion by U-Net; and (3) postprocessing. We detail each stage in the following subsections. [1]

A. PREPROCESSING

The proposed method takes a binary (black and white) image captured by a flat-bed scanner with a resolution of 150dpi as an input document image. Whereas many conventional OCR software typically requests higher resolution images (approximately 600 dpi) to prevent recognition errors, the proposed method can extract MEs from low-resolution images. This property also contributes to the efficiency of memory and computation time for subsequent ME detection processes using U-Net. If only a grayscale or color image is available, binarization with a threshold is requested.

To handle the white pixel regions that belong to the foreground, the input image is negated so that the characters and

[1]The implementation of the proposed method is available at https://github.com/uchidalab/MathExtraction_Unet

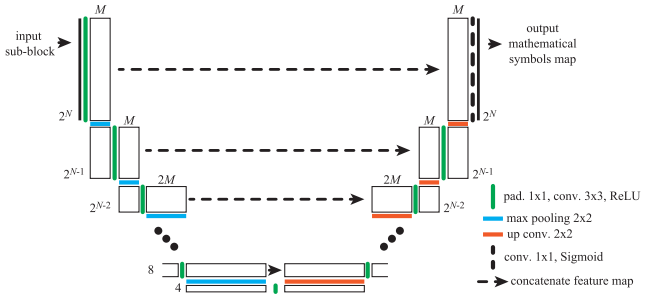


FIGURE 4. The U-Net architecture in the proposed method.

symbols are white. To prevent the elimination of thin components in the document image, white regions are dilated by 1 pixel in each direction using the mathematical morphology operation.

Overlapped square sub-blocks are defined to cover entire image regions and extracted to be the inputs to the image conversion stage. The edges of the input document image are wrapped by the opposite side of the image when the sub-block is over the image edges. The sub-block operation is used initially to deal with the memory size limits [7]. Also, the sub-block operation plays a role as data augmentation without image deformation operations. Generally, ideal estimation of variations of data is crucial for designing data augmentation protocols. The proposed method assumes that a flat-bed scanner captures the input images so that the input images do not contain significant image deformation. Consequently, data augmentation with such nonlinear image deformation is not required.

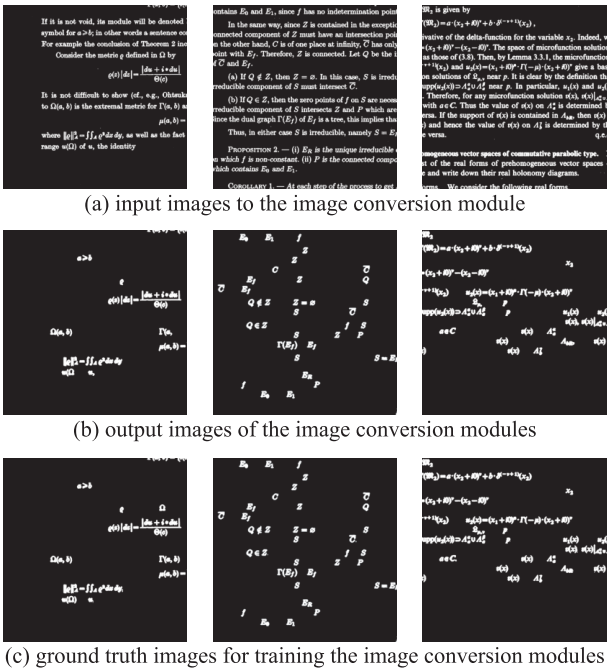
The width and height of the sub-blocks are parameters that affect the performance of the proposed method. The actual size of the sub-block images in our implementation was determined by the results of a preliminary experiment discussed in IV-C

B. ME DETECTION USING U-NET

ME detection in the proposed method can be considered as an image conversion task. Figure 5 shows examples of input, output and ground truth images of the ME detection process. As shown in the figure, the ME detection process is required to eliminate regions from MEs and extract the CCs that construct MEs.

We use the U-Net architecture proposed by Ronneberger *et al.* [7], motivated by the promising achievement of its semantic segmentation of biomedical images. U-Net is an FCN architecture that was proposed for the segmentation of biomedical images. By introducing skip connections between corresponding layers in the encoder and decoder, it successfully preserves the high-frequency components in the converted output images.

Figure 4 shows the actual U-Net configuration in the proposed method. The network mainly consists of two stages, i.e encoding and decoding stages. In encoding stage, the typical CNN architecture is employed. The encoding stage consists of multiple applications of a 3×3 convolution with a



(a) input images to the image conversion module

(b) output images of the image conversion modules

(c) ground truth images for training the image conversion modules

FIGURE 5. Examples of input, output and ground truth images for image conversion using U-Net.

1×1 padding followed by a rectified linear unit (ReLU) activate function and a 2×2 max-pooling operation for down sampling. The number of feature maps is doubled at each two downsampling steps. The decoding stages consists of an upsampling of the feature map followed by a 2x2 up-convolution. While the concatenation of feature maps in the original U-Net requires the cropping operation because there is loss of border pixels in every convolution, the proposed method does not employ cropping because the overlapped sub-blocks can recover the loss to each other. The final layer employs a 1x1 convolution to map each M-component feature vector to a binary output image.

As shown by Figure 4, the proposed method assumes that the size of an input sub-block is determined by $2^N \times 2^N$. The number of layers in the encoder and decoder stages is corresponding to the sub-block size. The base number of feature maps $M = 64$ is determined from the original U-Net impremention.

We implemented and trained U-Net to convert the input sub-block image to an image that contained only the CCs that constructed MEs. To achieve this conversion, we created ground truth images using an annotated dataset. In the dataset, we determined whether each character was a mathematical symbol or ordinary character. We eliminated the CCs annotated as ordinary characters to create the ground truth images. The ground truth in the training dataset was a set of sub-block images that were extracted at the corresponding position on the input image. We used the Dice loss determined by the following as the objective loss function to be minimized because the task of U-Net is binary-to-binary image conversion:

$$L(X, Y) = 1 - D(X, Y) = 1 - \frac{2|X \cap Y|}{|X| + |Y|}, \qquad (1)$$

where X and Y are the binary image of the network output and that of the ground truth. $X \cap Y$ denotes the overlap between X and Y, and $|X|$ is the L_1-norm of image X.

The proposed method does not use any information from either mathematical grammar or the character recognition results. The image conversion module in the proposed method is requested to obtain information that is crucial to determine the components that should remain as MEs only from the appearance of documents in the surrounding image area. A limited size of small regions may cause difficulty regarding making a decision even for humans in this scenario.

C. POSTPROCESSING

Through the image conversion process, we obtain sub-block images that contain CCs that correspond to MEs. In the postprocessing stage, we reconstruct the page image and extract CCs that correspond to mathematical symbols and characters.

To reconstruct the entire page image, each sub-block image is rearranged in the equivalent position and the maximum pixel value among the overlapping pixels is assigned to the corresponding pixel in the page image.

Additionally, pixel-wise multiplication between the resized reconstructed image and the original image is performed to eliminate dilated pixels caused by the morphology operation in preprocessing, and artifacts and noise injected during the image conversion process.

IV. PERFORMANCE EVALUATION

A. DATASETS

For a quantitative evaluation of the performance of mathematical OCR, a number of datasets have been proposed in the literature. InftyCDB datasets [39], [40] are large collections of mathematical symbols and notation from actual mathematical documents. UW databases [41] contain 100 typeset MEs from 25 document pages. However, these datasets are not applicable for evaluating ME detection performance because the content in the dataset is rearranged not to keep the original articles because of copyright reasons.

We collected two large datasets to train U-Net and evaluate the performance of the proposed ME detection method. The datasets, called GTDB-1 and GTDB-2, consist of document page images collected from scientific journals and textbooks. The GTDB-1 dataset, which was used to train the U-Net model, contains 31 English articles on mathematics. The GTDB-2 dataset, which was used for quantitative and qualitative evaluations of the performance of the proposed method, contains 16 articles. Diverse font faces and mathematical notation styles are included in these articles. A list of the articles in both datasets is provided in the appendix.

The statistics of each dataset are shown in Table 1. The article pages were originally scanned at 600 dpi. The ground truth annotations for each math symbol and ordinary character were attached manually. [2]

TABLE 1. Statistics of the datasets: Two datasets collected from scientific journals and textbooks.

	GTDB-1	GTDB-2
# articles	31	16
# pages	544	343
# math symbols	162,406	115,433
# of ordinary text characters	646,714	507,412

B. EVALUATION EXPERIMENTS

To train the U-Net model, we extracted 1,000 pairs of sub-blocks from each document page and the corresponding ground truth image from the GTDB-1 dataset. The locations of sub-blocks on each page image were determined randomly. The dataset consisted of 544 images, so the total number of sub-block images for training was 544,000.

We mainly used mathematical symbol (character) recall R_s, precision P_s and F-measure F_s as the performance measures. Each measure is defined as follows:

$$R_s = \frac{n_{TP}}{n_{TP} + n_{FN}}, \tag{2}$$

$$P_s = \frac{n_{TP}}{n_{TP} + n_{FP}}, \tag{3}$$

$$F_s = \frac{2\,P_s R_s}{P_s + R_s}, \tag{4}$$

where n_{TP}, n_{FP} and n_{FN} are the numbers of correctly detected mathematical symbols, falsely detected symbols or ordinary text, and undetected mathematical symbols, respectively. Pixel-level majority voting is used for the symbol-level evaluation. If the majority of pixels in a candidate symbol were detected as a mathematical symbol, the candidate symbol is determined as a mathematical symbol.

We also used ME-based recall (R_e), precision (P_e) and F-measure (F_e) as supplemental performance measures. Their definitions are similar to (2)–(4), but the numbers in the equations are counted for regions.

To determine MEs over the detected mathematical symbols, mathematical layout analysis is requested to obtain the spatial relationship between the symbols. We do not intend to implement layout analysis in the present study. Therefore, note that the evaluations using ME-based measures are based on the assumption that the candidates of mathematical regions are obtained using some layout analysis method. In this study, we extracted candidate regions using the ground truth so that the candidate regions contain in-line and displayed MEs, and

[2]The ground truth annotation has been released to the public to benchmark OCR performance for scientific documents. Although we could not include the original document images of articles for copyright reasons, we provide hyperlinks on our website to the web pages of the original documents, where the readers can obtain the document images: https://github.com/uchidalab/GTDB-Dataset/tree/master

ordinary text words. When the majority of symbols (characters) in a candidate region were detected as math symbols, the region was detected as an ME.

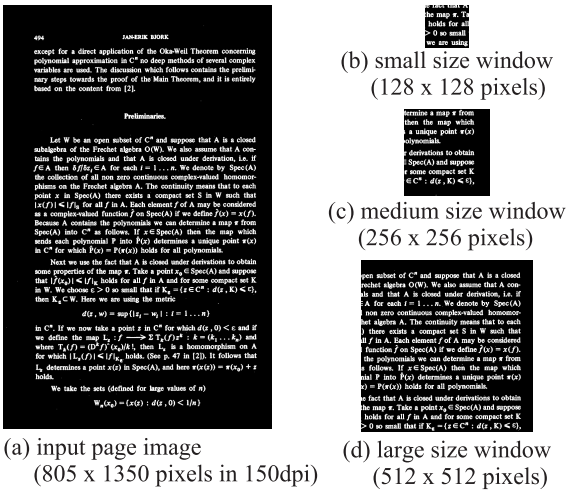
We used InftyReader [8] version 3.1.5.2 as the baseline for performance comparison. InftyReader is not only a public OCR software developed for documents containing MEs but is also recognized as a research achievement having the state-of-the-art performance for extraction and recognition of mathematical expressions. InftyReader uses the standard process pipeline shown in Figure 2. To prevent performance degradation, InftyReader requires input document images to be scanned at 600 dpi. To fair comparison, the same candidate regions are used for InftyReader.

C. PRELIMINARY EXPERIMENT FOR SUB-BLOCK SIZE SETTING

To determine the size of the sub-block images, we conducted a preliminary experiment. As described in III-B, the size of the sub-block images input to U-Net is an important parameter. It determines not only the actual configuration of U-Net but also the amount of information the network captures from the surrounding image regions.

In the preliminary experiment, small (128 × 128 pixels), medium (256 × 256) and large (512 × 512) sub-block sizes were tested, as shown in Figure 6. Each size of sub-block covered the image area of approximately 4.5 (small), 9.0 (medium) and 18.0 (large) text lines, respectively. One-document-out cross-validation was conducted for the test, which is a repeated procedure where document pages of one document article from the training dataset are reserved for testing and the remaining pages are used for training the image conversion module. The dataset consisted of 31 articles; therefore, the procedure was repeated 31 times.

Table 2 shows the results of the preliminary experiment. In the table, mean and standard deviation values of each performance measure are shown. The highest recall value was obtained by the small sub-block; however, the extracted results contained many false positives that should have belonged to ordinary text. In fact, many ordinary characters in italic or boldface were extracted as mathematical symbols. By contrast, the large sub-block successfully eliminated these false positives and provided the highest precision and F-measure values. Based on this result, we decided to use the large sub-block in our implementation.



(a) input page image (805 x 1350 pixels in 150dpi)
(b) small size window (128 x 128 pixels)
(c) medium size window (256 x 256 pixels)
(d) large size window (512 x 512 pixels)

FIGURE 6. The considered three sub-block sizes. The small, medium and large sub-blocks cover the area for which the height and width are approximately 4.5, 9.0 and 18.0 text lines, respectively.

D. PRELIMINARY EXPERIMENT FOR PRE- AND POST-PROCESSING

In the proposed method, we employ pre- and post-processing to mainly improve the stability of the proposed method for ME extraction. The dilation operation in the pre-processing is expected to enhance the recall and precision due to preventing the elimination of thin components in the document image. The pixel-wise AND operation between the output of U-net and the original image can enhance the precision due to prevent artifacts and noise.

We conducted a preliminary experiment to confirm the effectiveness of the pre- and post-process. In the preliminary experiment, we evaluated the ME detection performance in the case where the dilation operation in pre-processing and the pixel-wise AND operation in the post-processing are separately eliminated. Same as in the previous section, One-document-out cross-validation was conducted for each setup.

Table 3 shows the results of the experiment. As shown in the results, both operations contribute to improving mathematical symbol detection performance. Notably, the dilation process improves performance significantly. These results suggest that preventing the elimination of thin and small components in the document image is essential for the image conversion by U-net.

TABLE 2. Validating the performance of the proposed method for mathematical symbol detection against the size of sub-block images. Mean and standard deviation values of each performance measure are shown in the table. The large sub-block achieved the highest F measure value. Underlining in each column indicates the highest value of each measure.

	small	medium	large
R_s	0.927 ± 0.025	$\underline{0.956 \pm 0.011}$	0.952 ± 0.0092
P_s	0.791 ± 0.073	0.920 ± 0.020	$\underline{0.944 \pm 0.025}$
F_s	0.851 ± 0.034	0.937 ± 0.010	$\underline{0.947 \pm 0.016}$

TABLE 3. Validating the performance of the proposed method for mathematical symbol detection with and without the dilation and the pixel-wise AND operations. Mean and standard deviation values of each performance measure are shown in the table.

	w.o. dilation	w.o. AND	with both
R_s	0.798 ± 0.070	0.940 ± 0.010	$\underline{0.952 \pm 0.0092}$
P_s	0.639 ± 0.117	0.928 ± 0.022	$\underline{0.944 \pm 0.025}$
F_s	0.701 ± 0.049	0.933 ± 0.015	$\underline{0.947 \pm 0.016}$

TABLE 4. Performance comparison between the proposed method and InftyReader. The proposed method detected mathematical symbols and equations with higher performance to that of InftyReader. The right column shows the performance values after retraining U-Net. We discuss the retraining U-Net in subsection VI-B.

		InftyReader	proposed	after retraining
symbol detection	R_s	0.946	0.950	0.972
	P_s	0.971	0.973	0.974
	F_s	0.958	0.961	0.973
ME detection	R_e	0.950	0.952	0.970
	P_e	0.905	0.910	0.915
	F_e	0.928	0.931	0.941

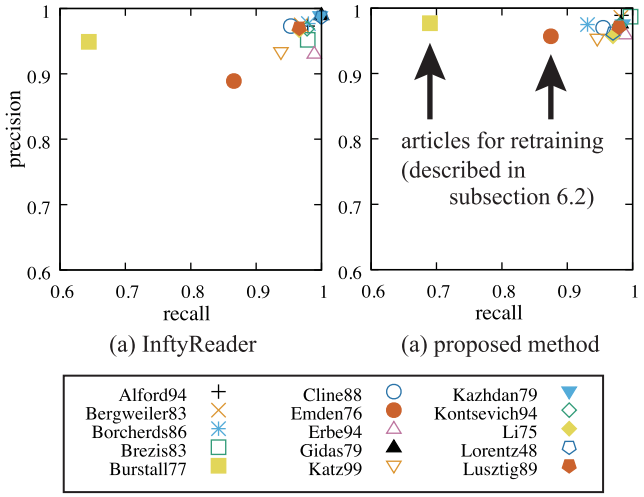


FIGURE 7. Performance comparison between the proposed method (b) and InftyReader (a). The plots show the mean values of symbol-based recall and precision for each article in the test dataset. We determined two articles as the candidates for retraining discussed in VI-B.

V. EXPERIMENTAL RESULTS

A. QUANTITATIVE EVALUATION OF ME DETECTION PERFORMANCE

Table 4 shows the comparison of detection performance between the proposed and baseline methods. The performance of the proposed method for both mathematical symbol and ME detection was slightly higher than that of the baseline method.

Figure 7 shows the symbol-based recall R_s and precision P_s for each article in the GTDB-2 test dataset obtained by InftyReader (a) and the proposed method (b). The performance of the proposed method and InftyReader was equivalently high for most of the articles. However, for two articles, Burstall and Emden, the precision value of the proposed method was much higher than that of InftyReader.

B. QUALITATIVE ANALYSIS OF THE DETECTION RESULTS

Figure 8 shows examples of mathematical symbol detection using the proposed method. The subfigures correspond to the page images of different articles. In the each figure, the cyan, magenta and yellow components indicate the correctly detected mathematical components (true positive: TP), falsely detected components (false positive: FP) and falsely undetected mathematical components (false negative: FN), respectively. Figure 8 (a)–(f) show the evaluation measures calculated for the page.

Figure 8(a) shows the result of mathematical symbol detection for a standard single column article. The proposed method successfully detected almost all mathematical components with very few FPs because the training dataset abundantly contained document pages in the standard one-column style.

Figure 8(b) is from an article with a double column style and small font. Although the training dataset used in the experiment did not contain any articles with a double column style, the proposed method could detect most of mathematical symbols. However, we observe that there were some FNs, mostly for in-line MEs.

Figure 8(c) was the worst case of detection in this experiment because the article used significantly different notation for MEs, in which some common English words (*trees, arrays*) in italic face were used as variables in MEs. Because some articles in the training dataset used italic face in paragraphs of ordinary text, the proposed method that was trained on such a dataset did not detect this notation. This was a difficult scenario, not only for the proposed method but also the baseline method. For the article, InftyReader could not detect most of these MEs either.

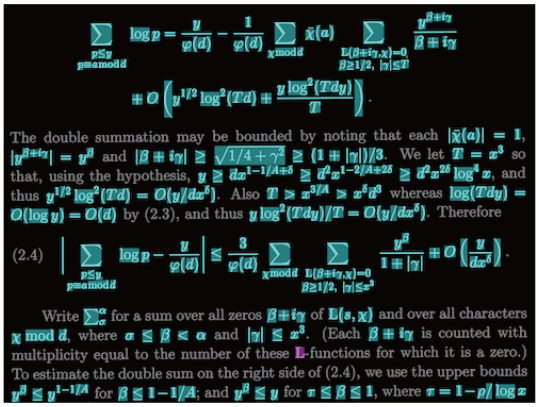
Figure 8(d) contains several mathematical characters in calligraphic face. Additionally, both Figs. 8(e) and (d) had old page styles, where spaces between symbols in MEs and line breaks between text lines were relatively wider than those of other articles. We observe that the proposed method was sufficiently robust for these variations because such variations also appeared in the training dataset.

Figure 9 illustrates the examples of the detection results of each method for the two articles, *Burstall* and *Emden*, mentioned in the previous subsection and Figure 7. These articles have typical styles, which distinguishes them from other articles. *Burstall* uses many ordinary texts in italic face. By contrast, *Emden* uses a number of bold roman face in both in-line and displayed MEs. Because such mathematical symbols were rare in the training dataset, it was quite difficult for the proposed method to detect these symbols in the mathematical equations.

VI. DISCUSSION

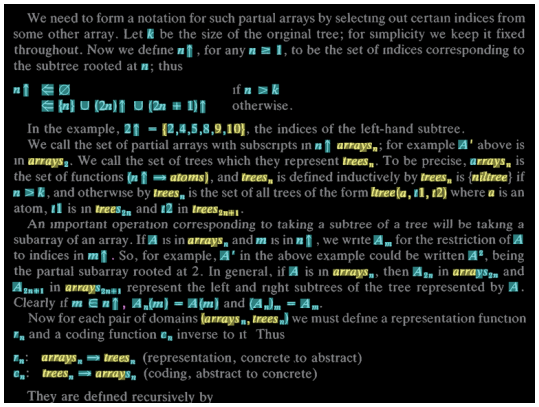
A. PERFORMANCE VARIATION OVER THE TRAINING DATA SIZE

In the experiment, we collected a massive training dataset that contained 544 pages with annotations. The implemented U-Net framework has the ability to convert document images from such a massive dataset, but we should consider the size of the training dataset. If a smaller dataset contributes to higher performance, then we could reduce the labor for the annotation task and time for training the framework.

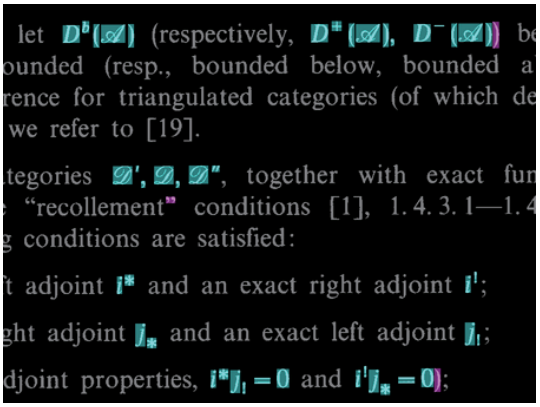


(a) *R*s=0.995, *P*s=0.998, *F*s=0.997

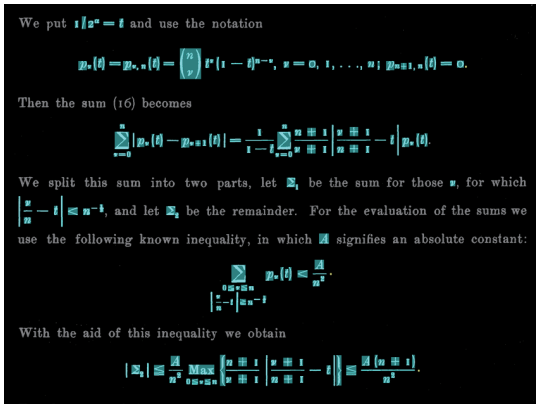
(b) *R*s=0.978, *P*s=0.985, *F*s=0.981



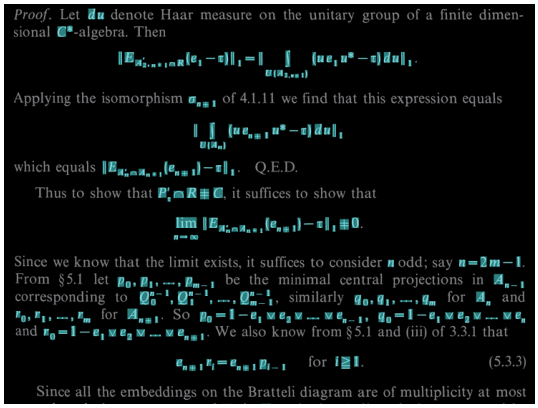
(c) *R*s=0.554, *P*s=0.957, *F*s=0.702



(d) *R*s=0.964, *P*s=0.912, *F*s=0.937



(e) *R*s=0.995, *P*s=0.996, *F*s=0.997



(f) *R*s=1.000, *P*s=0.971, *F*s=0.985

FIGURE 8. Examples of mathematical symbol and expression detection results for the proposed method. (a)–(f) show cropped regions from page images in different articles. The components in cyan, magenta and yellow denote the TP, FP and FN components, respectively. The sub-captions show the recall, precision and *F*-measure values for each page.

We created two types of subsets of the training dataset that consisted of a fixed number of document page images. The document page images in the first subset were collected in an article-wise manner from the full dataset. This means that all page images in the determined articles were included in the subset. The second subset was a collection of the same number of page images that were randomly selected from the full dataset. We expect that a diversity of font faces and mathematical notation styles is kept by the random selection. We adjusted the number of images in the subsets and repeated the experiment for U-Net training and evaluation.

Figure 10 shows the results of the experiments. The vertical and horizontal axes denote the symbol-based F-measure value F_s and number of pages in the subset, respectively. Whereas the F-measure value rapidly decreased when the number of pages in the article-based subset

(a) Burstall : InftyReader (b) Burstall : proposed

(c) Emden : InftyReader (d) Emden : proposed

FIGURE 9. **Visual assessment of the detected results for the proposed method and InftyReader. (a) and (b) are examples of the detected results for a page from *Burstall*; (c) and (d) are those for a page from *Emden*.**

was less than 50, the randomly selected subset retained the high F-measure value.

This result suggests that the number of page images can be reduced if a diversity of documents and mathematical styles is ensured. Furthermore, when the total page number in the training dataset is limited, random selection is more effective to maintain performance.

B. RETRAINING THE PROPOSED METHOD

The proposed method tended to fail to extract mathematical symbols and expressions from document page images for which document styles were not covered by the training dataset, as we showed in the previous subsection. Such scenarios may occur frequently in actual applications. One way to manage such scenarios is to adopt retraining using an updated training dataset.

To evaluate the adaptivity of the proposed method to the retraining scenario, we conducted a retraining experiment that aimed to improve the detection performance for the two articles we discussed in the previous subsection. The retraining scenario can be seen as a fine-tuning with a retraining dataset which consists of the original training page images

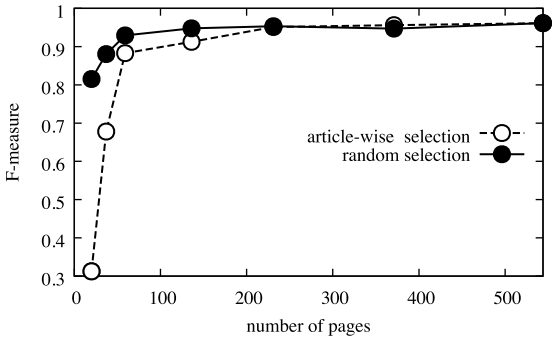


FIGURE 10. Variation of the *F*-measure value with the number of document pages contained in the training dataset. Article-wise selection means that the document pages used for training were selected according to each article. Random selection means that the document pages were selected randomly from all document pages. Random selection achieved higher *F*-measure values when the number of pages in the training dataset was small.

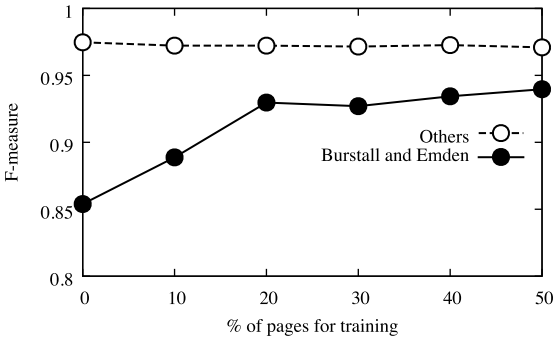
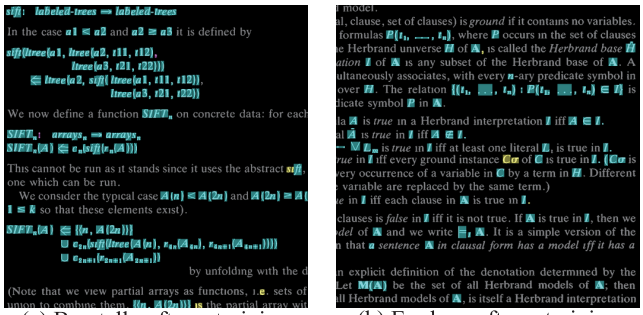


FIGURE 11. Effects of retraining with the dataset including a part of the target article. The vertical and horizontal axes are the *F*-measure value and number of injected pages, respectively. The plots titled "Burstall and Emden", and "Others" denote the *F*-measure values of retrained and remaining articles, respectively. The Burstall and Emden articles consist of 24 and 10 pages, respectively.

and the injected page images from target article. We selected two articles, Burstall and Emden, as the target articles from the GTDB-2 dataset. We extracted parts of the pages from each article and injected them into the training dataset. From the results in VI-A, we knew that the size of the training dataset could be reduced. Considering the size of the injected pages from the target articles, we used 50 randomly selected pages from the training dataset for retraining. We evaluated the improvement of the symbol-based F-measure F_s for the remaining pages of the target articles and the remaining articles in the test dataset against the size of the injected target article pages. The injected page images were eliminated from the evaluation dataset.

Figure 11 shows the results of the retraining experiment. The vertical and horizontal axes denote the symbol-based F-measure value F_s and the number of injected target article pages, respectively. The F-measure value of the target articles improved with the increase of the size of the injected article pages. By contrast, whereas the F-measure value of the other articles slightly decreased, the total F-measure value maintained an almost constant level. The right column in Table 4 shows the performance values after the retraining using 50% of pages from the target articles. All evaluation values are



(a) Burstall : after retraining (b) Emden : after retraining

FIGURE 12. Examples of the detection results after retraining. (a) and (b) correspond to Figure 9 (b) and (d). The detection results improved using retraining. Note that these page images were not included the additional dataset for retraining.

successfully improved by the retraining. Figure 12 illustrates examples of the detection results after retraining. We observe that the many errors contained in Figs. 9 (b) and (d) were improved by retraining.

These results suggest the following: When the proposed method encounters articles that have new or special document styles and/or mathematical notation, the method may not extract MEs appropriately from such articles at that moment. However, once users provide additional annotations for some pages from the target article, detection performance can be improved by retraining. The ease of retraining is a superiority of the proposed method compared with conventional rule-based methods.

VII. CONCLUSION

In this paper, we proposed a mathematical symbol and expression detection method that uses image conversion by a U-Net framework. By using image-to-image conversion using U-Net, the proposed method extracts mathematical symbols and expressions without any information from mathematical and linguistic grammar. Consequently, the proposed method can bypass the standard process pipeline of mathematical equation recognition.

The detection performance of the proposed method was evaluated using a large page image dataset collected from real scientific journals and textbooks. The experimental results confirm that the detection accuracy is superior to that of the baseline InftyReader. Additionally, the proposed method is applicable to the size reduction of a training dataset and retraining using an injected training dataset.

The dataset used in the performance evaluation has been released to the public for benchmarking OCR performance for scientific documents. For the details of the dataset, please see the appendix or the dataset website.

Future research topics will involve (1) integrating the proposed ME detection into the mathematical OCR pipeline; (2) the adaptation and integration of scientific documents in other languages; and (3) the application of the image conversion framework for the detection of other document page elements, that is, figures and tables.

APPENDIX

LIST OF ARTICLES IN GTDB-1 DATASET

Acta Math., 124, 2, 37-63, 1970; *ibid.*, 181, 2, 283-305, 1998; Ann. Inst. Fourier, 20, 1, 493-498, 1970; *ibid.*, 49, 2, 375-404, 1999; *ibid.*, 49, 2, 375-404, 1999; Analytical Mechanics (pp. 493-557); Cambridge Univ. Press, 1998; Ann. Math., 91, 550-569, 1970; Ann. Math. Studies, 66, 157-173, 1971; Arkiv für Matematik, 9(1), 141-163 1971; Ark. Mat., 35, 185-199, 1997; Ann. Sci. École Norm. Sup., 4d sér, t.3, 273-284, 1970; *ibid.*, 4e sér, t.30, 367-384, 1997; Bull. Amer. Math. Soc., 77(1), 157-159, 1971; *ibid.*, 80(6), 1219-1222,1974; *ibid.*, 35(2), 123-143, 1998; Bull. Soc. Math. France, 98, 165-192, 1970; *ibid.*, 126, 245-271, 1998; Invent. Math., 9, 121-134, 1970; *ibid.*, 138, 163-181, 1999; J. Math. Soc. Japan, 27(2), 281- 288, 1975; *ibid.*, 27(2), 289-293, 1975; *ibid.*, 27(2), 497-506, 1975; J. Math. Kyoto Univ., 11(1), 181-194, 1971; *ibid.*, 11(1), 373-375, 1971; *ibid.*, 11(2), 377-379, 1971; Kyushu J. Math., 53, 17-36, 1999; Math. Ann., 225(3), 275-292, 1977; *ibid.*, 315, 175-196, 1999; Tohoku Math. J., 25, 317-331, 1973; *ibid.*, 25, 333-338, 1973; *ibid.*, 42, 163-193, 1990.

LIST OF ARTICLES IN GTDB-2 DATASET

Alford94:Ann. Math., 140, 703-722, 1994; *Bergweiler83*:Bull. Amer. Math. Soc., 28(2), 151-188, 1993; *Borcherds86*:Proc. Natl. Acad. Sci. USA, 83, 3068-3071, 1986; *Brezis83*:Proc. Amer. Math. Soc., 88(3) 486-490, 1983; *Burstall77*:J. Assoc. Comp. Mach., 24(1) 44-67, 1977; *Cline88*:j. reine angew. Math. 391, 85-99, 1988; *Emden76*:J. Assoc. Comp. Mach., 23(4) 733-742, 1976; *Erbe94*:Proc. Amer. Math. Soc., 120(3) 743-748, 1994; *Gidas79*:Commun. Math. Phys. 68, 209-243, 1979; *Jones83*:Invent. math. 72, 1-25, 1983; *Katz99*:Bull. Amer. Math. Soc., 36(1), 1-26, 1999; *Kazhdan79*:Invent. math. 53, 165-184, 1979; *Kontsevich94*:Commun. Math. Phys. 164(3), 525-562, 1994; *Li75*:Amer. Math. Mon., 82(10), 985-992, 1975; *Lorentz48*:Acta Math., 80(1), 167-190, 1948; *Lusztig89*:J. Amer. Math. Soc., 2(3), 599-635, 1989.

WEBSITE OF THE DATASETS

https://github.com/uchidalab/GTDB-Dataset/tree/master

REFERENCES

[1] R. Zanibbi and D. Blostein, "Recognition and retrieval of mathematical expressions," *Int. J. Document Anal. Recognit.*, vol. 15, no. 4, pp. 331–357, Dec. 2012.

[2] Y. Deng, A. Kanervisto, J. Ling, and A. M. Rush, "Image-to-markup generation with coarse-to-fine attention," in *Proc. 34th Int. Conf. Mach. Learn.*, Aug. 2017, pp. 980–989.

[3] J. Zhang, J. Du, and L. Dai, "A GRU-based encoder-decoder approach with attention for online handwritten mathematical expression recognition," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 902–907.

[4] J. Zhang, J. Du, S. Zhang, D. Liu, Y. Hu, J. Hu, S. Wei, and L. Dai, "Watch, attend and parse: An end-to-end neural network based approach to handwritten mathematical expression recognition," *Pattern Recognit.*, vol. 71, pp. 196–206, Nov. 2017.

[5] J. Zhang, J. Du, and L. Dai, "Track, attend, and parse (TAP): An end-to-end framework for online handwritten mathematical expression recognition," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 221–233, Jan. 2019.

[6] W. He, Y. Luo, F. Yin, H. Hu, J. Han, E. Ding, and C.-L. Liu, "Context-aware mathematical expression recognition: An end-to-end framework and a benchmark," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 3246–3251.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention* (Lecture Notes in Computer Science), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[8] M. Suzuki, F. Tamari, R. Fukuda, S. Uchida, and T. Kanahori, "INFTY: An integrated OCR system for mathematical documents," in *Proc. ACM Symp. Document Eng.*, Nov. 2003, pp. 95–104.

[9] U. Garain, "Identification of mathematical expressions in document images," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, Jul. 2009, pp. 1340–1344.

[10] S. Yamazaki, F. Furukori, Q. Zhao, K. Shirai, and M. Okamoto, "Embedding a mathematical OCR module into OCRopus," in *Proc. 11th Int. Conf. Document Anal. Recognit.*, Sep. 2011, pp. 880–884.

[11] U. Garain and B. B. Chaudhuri, "OCR of printed mathematical expressions," in *Digital Document Processing* (Advances in Pattern Recognition). London, U.K.: Springer, 2007, pp. 235–259.

[12] Y.-C. Lin, C.-Y. Wang, and J.-Y. Zeng, "A case study on mathematical expression recognition to GPU," *J. Supercomput.*, vol. 73, no. 8, pp. 3333–3343, Aug. 2017.

[13] B. B. Chaudhuri and U. Garain, "An approach for recognition and interpretation of mathematical expressions in printed document," *Pattern Anal. Appl.*, vol. 3, no. 2, pp. 120–131, Jun. 2000. doi: 10.1007/s100440070017.

[14] *Mathpix*. Accessed: Jul. 15, 2019. [Online]. Available: https://mathpix.com/

[15] H.-J. Lee and J.-S. Wang, "Design of a mathematical expression understanding system," *Pattern Recognit. Lett.*, vol. 18, no. 3, pp. 289–298, 1997.

[16] A. Kacem, A. Belaïd, and M. B. Ahmed, "Automatic extraction of printed mathematical formulas using fuzzy logic and propagation of context," *Int. J. Document Anal. Recognit.*, vol. 4, no. 2, pp. 97–108, Dec. 2001.

[17] A. K. Das, S. P. Chowdhury, S. Mandal, and B. Chanda, "Automated segmentation of math-zones from document images," in *Proc. 7th Int. Conf. Document Anal. Recognit.*, vol. 2, Aug. 2003, pp. 755–759.

[18] T.-Y. Chang, Y. Takiguchi, and M. Okada, "Physical structure segmentation with projection profile for mathematic formulae and graphics in academic paper images," in *Proc. 9th Int. Conf. Document Anal. Recognit.*, vol. 2, Sep. 2007, pp. 1193–1197.

[19] D. M. Drake and H. S. Baird, "Distinguishing mathematics notation from english text using computational geometry," in *Proc. 8th Int. Conf. Document Anal. Recognit.*, vol. 2, Aug./Sep. 2005, pp. 1270–1274.

[20] B. H. Phong, T. M. Hoang, and T.-L. Le, "A new method for displayed mathematical expression detection based on FFT and SVM," in *Proc. 4th NAFOSTED Conf. Inf. Comput. Sci.*, Nov. 2017, pp. 90–95.

[21] B. H. Phong, T. M. Hoang, and T.-L. Le, "Mathematical variable detection based on convolutional neural network and support vector machine," in *Proc. Int. Conf. Multimedia Anal. Pattern Recognit. (MAPR)*, May 2019, pp. 1–5.

[22] U. Garain, B. B. Chaudhuri, and A. R. Chaudhuri, "Identification of embedded mathematical expressions in scanned documents," in *Proc. 17th Int. Conf. Pattern Recognit.*, vol. 1, Aug. 2004, pp. 384–387.

[23] K.-F. Chan and D.-Y. Yeung, "Mathematical expression recognition: A survey," *Int. J. Document Anal. Recognit.*, vol. 3, no. 1, pp. 3–15, Aug. 2000.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," Sep. 2014, *arXiv:1409.0473*. [Online]. Available: https://arxiv.org/abs/1409.0473

[26] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[27] K. Cho, A. Courville, and Y. Bengio, "Describing multimedia content using attention-based encoder-decoder networks," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1875–1886, Nov. 2015.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[29] W. Zhang, Z. Bai, and Y. Zhu, "An improved approach based on CNN-RNNs for mathematical expression recognition," in *Proc. 4th Int.*

Conf. Multimedia Syst. Signal Process. (ICMSSP), New York, NY, USA, May 2019, pp. 57–61. doi: 10.1145/3330393.3330410.
[30] X. Lin, L. Gao, Z. Tang, J. Baker, and V. Sorge, "Mathematical formula identification and performance evaluation in PDF documents," *Int. J. Document Anal. Recognit.*, vol. 17, no. 3, pp. 239–255, Sep. 2014.
[31] L. Gao, X. Yi, Y. Liao, Z. Jiang, Z. Yan, and Z. Tang, "A deep learning-based formula detection method for PDF documents," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 553–558.
[32] A. D. Le and M. Nakagawa, "Training an end-to-end system for handwritten mathematical expression recognition by generated patterns," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 1056–1061.
[33] K. Ma, Z. Shu, X. Bai, J. Wang, and D. Samaras, "DocUNet: Document image unwarping via a stacked U-net," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4700–4709.
[34] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "ICDAR2017 competition on document image binarization (DIBCO 2017)," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 1395–1403.
[35] M. Fink, T. Layer, G. Mackenbrock, and M. Sprinzl, "Baseline detection in historical documents using convolutional U-nets," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst.*, Apr. 2018, pp. 37–42.
[36] G. Renton, C. Chatelain, S. Adam, C. Kermorvant, and T. Paquet, "Handwritten text line segmentation using fully convolutional network," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 5, Nov. 2017, pp. 5–9.
[37] S. Capobianco, L. Scommegna, and S. Marinai, "Historical handwritten document segmentation by using a weighted loss," in *Proc. IAPR Workshop Artif. Neural Netw. Pattern Recognit.*, 2018, pp. 395–406.
[38] S. A. Oliveira, B. Seguin, and F. Kaplan, "dhSegment: A generic deep-learning approach for document segmentation," Apr. 2018, *arXiv:1804.10371*. [Online]. Available: https://arxiv.org/abs/1804.10371
[39] M. Suzuki, S. Uchida, and A. Nomura, "A ground-truthed mathematical character and symbol image database," in *Proc. 8th Int. Conf. Document Anal. Recognit.*, vol. 2, Aug./Sep. 2005, pp. 675–679.
[40] S. Uchida, A. Nomura, and M. Suzuki, "Quantitative analysis of mathematical documents," *Int. J. Document Anal. Recognit.*, vol. 7, no. 4, pp. 211–218, 2005.
[41] I. T. Phillips, "Methodologies for using UW databases for OCR and image-understanding systems," *Proc. SPIE, Document Recognition V*, pp. 112–127, Apr. 1998.

WATARU OHYAMA received the B.E., M.E., and Ph.D. degrees in engineering from Mie University, in 1998, 2000, and 2007, respectively. From 2000 to 2018, he was an Assistant Professor with Mie University and an Associate Professor with Kyushu University. He is currently a Professor with the Saitama Institute of Technology. His research interests include pattern recognition and image processing. He is a member of IEICE, IEEJ, and IPSJ.

MASAKAZU SUZUKI received the B.Sc. and M.Sc. degrees from Kyoto University, in 1969 and 1971, respectively, and the D.d'Etaés Sc. degree from University of Paris VII, in 1977. During his career with CNRS, from 1975 to 1977, and with Kyushu University, from 1977 to 2010, where he is currently a Professor Emeritus with the Faculty of Mathematics. His main research interests include complex analysis, algebraic geometry, mathematical document recognition, and mathematical knowledge management. He is a member of MSJ, IEICE, JASE, and JALD.

SEIICHI UCHIDA received the B.E., M.E., and Dr.Eng. degrees from Kyushu University, in 1990, 1992, and 1999, respectively. From 1992 to 1996, he joined SECOM Company Ltd., Japan. He is currently a Distinguished Professor with Kyushu University. His research interests include pattern recognition and image processing. He is a member of IEICE and IPSJ. He was a recipient of the 2007 IAPR/ICDAR Best Paper Award and the 2010 ICFHR Best Paper Award.