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.

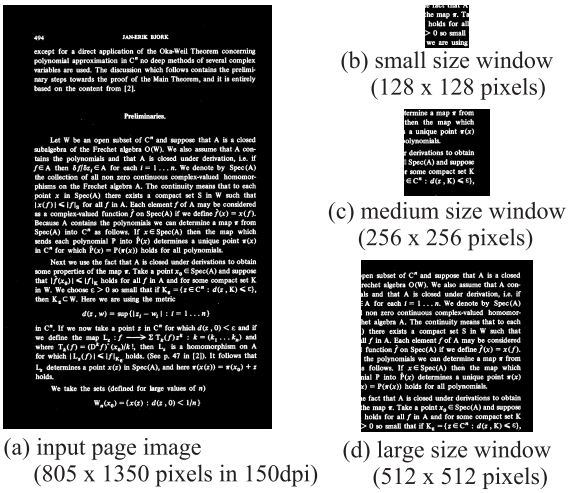
We used InftyReader [8] version 3.1.5.2 as the baseline for performance comparison. InftyReader is not only a public OCR software developed for documents containing MEs but is also recognized as a research achievement having the state-of-the-art performance for extraction and recognition of mathematical 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	$\underline{0.947 \pm 0.016}$