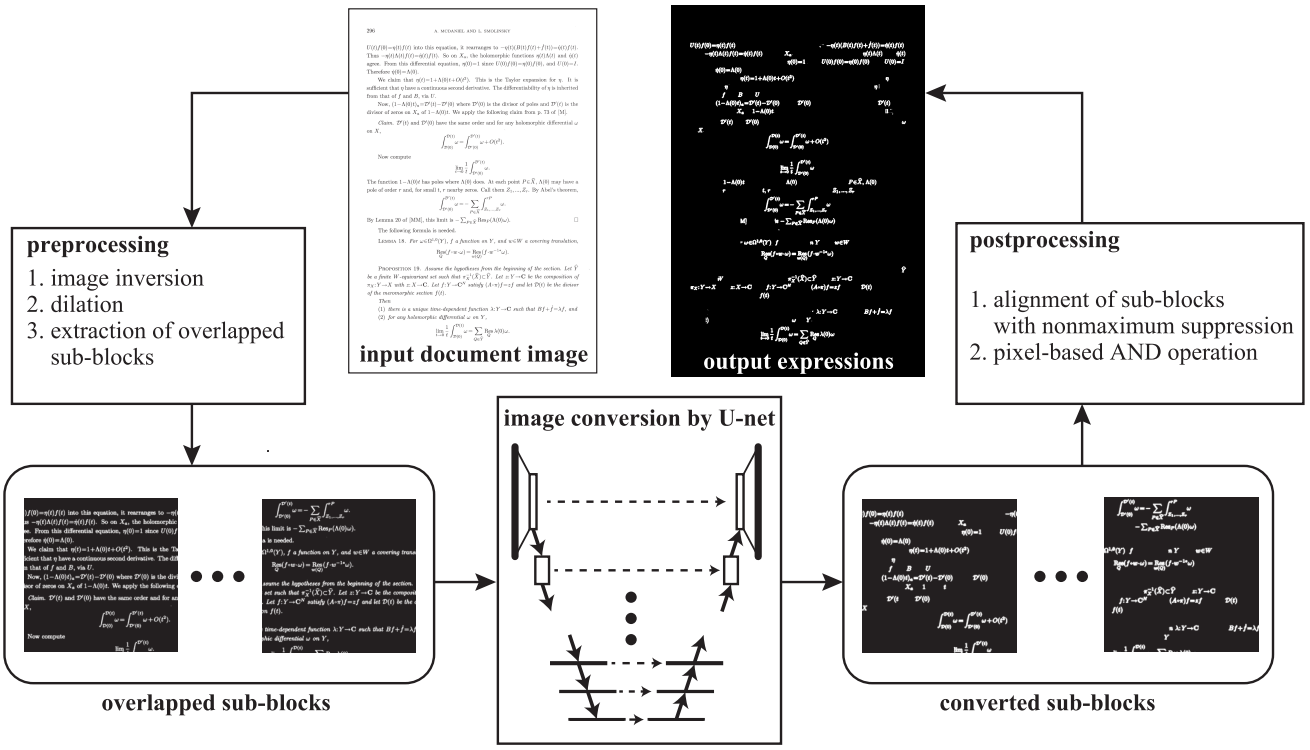

FIGURE 3. Outline of the proposed method. The method consists of three main stages: preprocessing, image conversion by U-Net and postprocessing.

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