Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

Implementation of Tesseract OCR and Bounding Box for Text Extraction on Food Nutrition Labels

The Manuel Eric Saputra[*], Ajib Susanto, Bastiaans Jessica Carmelita

Faculty of Computer Science, Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]manueleric.saputra@gmail.com, [2]ajib.susanto@dsn.dinus.ac.id, [3]jessicacarmelita2004@gmail.com
Correspondence Author Email: manueleric.saputra@gmail.com
Submitted: **21/10/2024**; Accepted: **01/12/2024**; Published**: 03/12/2024**

Abstract−This study focuses on implementing Optical Character Recognition (OCR) using the Tesseract engine, integrated with bounding box detection, to extract nutritional information from food nutrition labels. The research addresses the challenge of limited consumer access to and understanding of nutritional data, a factor contributing to health issues such as obesity and related metabolic disorders. Studies indicate that although Indonesian consumers generally have a good level of knowledge and positive attitudes toward nutritional labels, the actual behavior of reading and understanding these labels remains limited. Additionally, packaged foods consumed outside the home constitute a significant portion of daily caloric intake, which can lead to health complications if not properly managed. With obesity levels among adults in Indonesia rising to concerning rates, this study highlights the importance of providing accessible nutritional data. In this work, MobileNetV1 is used as the backbone model for bounding box detection, effectively identifying and isolating label regions to enhance OCR accuracy. Tesseract OCR, known for its LSTM-based architecture, is applied to predict sequential data patterns, such as rows of text on nutrition labels. Preprocessing techniques, including grayscale conversion, brightness adjustment, CLAHE (Contrast Limited Adaptive Histogram Equalization), and denoising, are used to improve text clarity and further refine OCR output accuracy. Post-processing steps involve rule-based and contextual error correction to handle common OCR inaccuracies. Evaluated on 10 different label images, the system achieved a maximum Word Error Rate (WER) of 10% and a Character Error Rate (CER) of 1.6%, demonstrating high accuracy in nutritional information extraction.

Keywords: OCR; Bounding Box; Preprocessing; Postprocessing, Nutrition Labels; Word Error Rate; Character Error Rate

1. INTRODUCTION

In today's modern era, public awareness of health is increasing, so nutritional information on food and beverage packaging products is an important factor for consumers to consider decision making. The average consumer in Indonesia has good knowledge and attitude towards nutritional value information labels (87.8%), but this is not matched by the behavior of reading nutritional value information (52.3%) [1], [2]. Based on a survey conducted by Yunita Diansa Sari, et al, food consumed outside the home (packaging) accounts for 34.4% of the total daily energy intake in Indonesia [3]. However, if packaged food or drinks are consumed in excess, it can cause negative impacts on health, especially because it has the potential to provide energy or calorie intake that exceeds the body's needs. This excess intake can trigger various health problems, such as weight gain, obesity, and can be a risk factor for diseases such as diabetes, heart disease, cancer and other metabolic disorders [3], [4], [5]. A study in the United States also showed that packaged foods, fast food and soft drinks accounted for more than a third of daily calorie needs leading to increased energy intake that exceeds needs and causes obesity. Based on the latest data from the Ministry of Health of the Republic of Indonesia in 2023 shows that the level of obesity in the population over the age of 18 reached an alarming figure of around 28%, a significant increase from previous years [5]. One of the main factors causing the problem is the lack of understanding of the nutritional information listed on the products consumed daily. Nutritional information on product packaging plays an important role in helping consumers choose products that meet their health needs. Nutrition labeling on product packaging can benefit a wide range of consumer characteristics [6], [7] but its effectiveness tends to be more significant for individuals with lower health awareness [6]. However, in some cases, there are still many individuals who have difficulty estimating and measuring the amount of food intake due to lack of nutritional information, manual process of writing down detailed nutritional information and other reasons [8]. Furthermore, not all consumers have the time or visual ability to effectively understand nutrition labels or they are not clearly visible [1], [5].

With technological developments in image processing, it is possible to automatically detect and extract nutritional information from product packaging. Researchers use Optical Character Recogniton (OCR) technology to extract text from images such as nutrition labels, notes, books, vehicle plates and so on. Tesseract, an OCR engine developed by Google, is often used by researchers because of its accuracy in extracting text from images [9], [10]. Previous researchers used Tesseract and Abby Fine Reader for text extraction from images to be processed with transformer models. Text extraction results using Tesseract show better accuracy than Abby Fine Reader, with 76% accuracy compared to 62% in the BERT model, and 87% compared to 78% in the RoBERTa model [11]. Other researchers have also used the Tesseract OCR engine for text extraction from nutrition label images, as the LSTM architecture used is capable of storing long-term information [12]. This makes it effective in understanding and predicting patterns in sequential data, such as rows of text on nutrition labels [12]. Tesseract is of course also used for text detection from images of product packaging ingredients [13]. In text extraction from book images, Tesseract performs better if the image has a grayscale or black and white scale [14]. The researchers also compared Tesseract and GoCR based on precision and accuracy using various parameters, such as image type, font type, brightness, and

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

resolution, and concluded that Tesseract is superior to GoCR in most cases [15]. Furthermore, among the open-source category, the average error rates of OCRopus and Tesseract are relatively similar, but OCRopus requires a longer conversion time, so Tesseract is considered a more efficient choice [15], [16]. Research also compared the performance of open-source Tesseract with Transym for ANPR, showing that Tesseract has higher accuracy for color and grayscale images, runs faster, and has a lower standard deviation of accuracy than Transym [15].

Based on the research that has been done, the author uses the Tesseract OCR engine in text extraction from images of packaged nutrition labels. Tesseract OCR has significant advantages due to its LSTM-based architecture, which is capable of storing long-term information and is effective in predicting sequential data patterns, such as text sequences on nutrition labels [12]. This, along with better accuracy, efficiency, and speed compared to other OCR models, makes it a superior choice for text extraction on images with various conditions, such as grayscale or color [11], [14], [15]. Tesseract is able to recognize different font types and text sizes, and remains accurate even in low-quality images. In addition, Tesseract can be trained to handle specialized text formats such as nutrition numbers and ingredient lists, making it effective and efficient for information extraction from food labels [13]. To support the accuracy of text extraction from nutrition label images, the authors apply the MobileNetV1 model as an object detection method specifically used to identify and extract nutrition labels from images. The main advantage of MobileNetV1 in object detection is that it has high efficiency, as it uses depth-wise separable convolution which reduces the computational complexity significantly [19]. This allows MobileNetV1 to be used on devices with low processing power, such as cell phones and other mobile devices. When MobileNetV1 is combined with the Single Shot Multibox Detector (SSD) algorithm, object detection can be performed in real-time at high speed without sacrificing accuracy [19]. This combination provides an optimal balance between speed and accuracy, making it ideal for applications that require fast performance with limited resources [20], [21], [22], [23].

This research aims to develop an accurate bounding box detection method and apply text detection and OCR with Tesseract to extract nutritional information from product packaging labels. The implementation of this technology is expected to make a significant contribution in increasing public awareness of nutrition, thus helping to reduce the prevalence of obesity in Indonesia through the provision of easily accessible and understandable nutritional information. The contribution of this research is validated with all stages of extraction with OCR on food nutrition labels with the stages of bounding box detection, image cropping, image preprocess, text extraction, image postprocess.

2. RESEARCH METHODOLOGY

2.1 Data Collection

Data collection in this study was carried out by taking photos of nutrition labels from various packaged food products. The photographing process was done manually using a smartphone with a JPG image format, which was chosen because its quality is good enough for information extraction purposes at a later stage. This study only involved food products, without including beverage products.

The images were taken under optimal lighting conditions to ensure that the information on the nutrition labels could be clearly read. The researcher tried to ensure that there were no shadows or distortions in the images, and ensured that the entire label was visible in one frame. Pictures were taken of various packaged food products on the market, including snacks, ready meals and other processed foods.

Figure 1. Nutrition Label Data Collection

Figure 1 displays the data collected for testing purposes, focusing on nutrition labels on food products, which provide essential information. This includes serving size, indicating the recommended portion; calorie count, representing the calories in a single serving; and fat content, detailing total fat along with saturated and trans fats.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

Additionally, the labels include total carbohydrate information, encompassing dietary fiber and sugar; sugar content, specifying the total sugar in the product; protein, indicating the protein content per serving; and sodium, providing information on salt content. The dataset created from these captured images will be utilized for further processing and analysis, specifically aimed at extracting nutrition information from a variety of food products.

2.2 Bounding Box

The author uses the MobileNetV1 model structure as a feature extractor that aims to detect bounding boxes on product packaging nutrition labels. MobileNetV1 can be used to detect or classify objects [24]. This model begins with an image preprocessing process which includes several steps, namely converting the input image in the form of a tensor into a float and normalizing the image. This preprocessing is done through the ResizeImage operation using the ResizeBilinear method, which ensures that the image has dimensions that match the model input for further processing. This process is very important because the input must match the format that has been set by the network architecture to keep it consistent and stable.

After preprocessing, the input image is then processed through the convolution layers in MobileNetV1. MobileNetV1 is a lightweight neural network model designed to run on devices with low computing power. MobileNetV1 uses the depthwise separable convolutions technique which is a combination of depthwise convolution and pointwise convolution, where Depthwise convolution processes each input channel separately, while Pointwise convolution combines the results of the depthwise convolution to produce a more comprehensive feature output. This technique reduces the number of parameters required and optimizes the performance of the model to make it lighter and faster. Each convolution layer is followed by a Batch Normalization layer that helps normalize the output of the previous layer, and ReLU6 activation that adds non-linearity to the model to enable detection of more complex features such as recognizing important patterns of the image i.e. shapes, textures, and lines used in the product packaging nutrition label detection process.

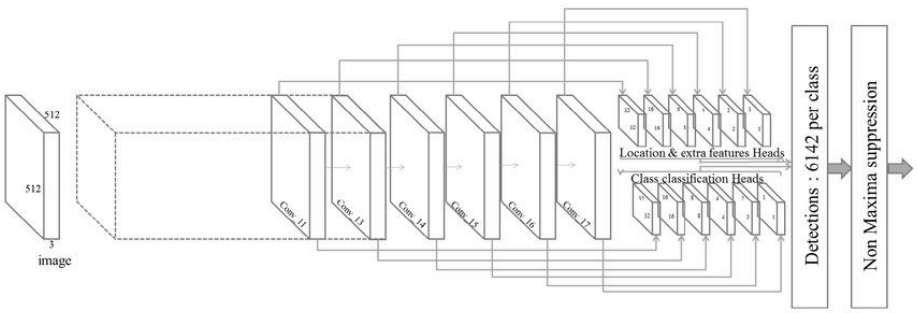


Figure 2. MobileNetV1 Architecture - SSD

As in Figure 3, MobileNetV1 also consists of several nested layers starting with Conv2d_0 to Conv2d_13 such as `Conv2d_1_depthwise` and `Conv2d_1_pointwise`, which serve to extract features from the input image at various levels. Each layer uses statistical computation through batch normalization layer, and is equipped with ReLU6 activation function to maintain the stability, efficiency and accuracy of feature recognition in model training. Depthwise convolutions allow convolution operations to be performed on each channel independently, while Pointwise convolutions (1x1 convolutions) are used to combine features from different channels of depthwise convolution results. The model has a detection component that predicts the bounding box for the detected objects in the image. MobileNetV1 serves as the feature extractor, and Single Shot Detector (SSD) performs object detection (localization) and classification [25]. SSD is a feed-forward convolution-based object detector that generates a collection of bounding boxes with their values and presents the class of each bounding box [24]. This bounding box detection model is capable of real-time nutrition label detection while maintaining high accuracy and has an optimal architecture for resource-constrained devices.

2.3 Preprocessing Data

Image preprocessing is an important step to improve the accuracy of text extraction using OCR (Optical Character Recognition). Nutrition table images in datasets often face challenges such as uneven lighting, low contrast, and noise. Therefore, various preprocessing techniques are applied to ensure the text can be clearly recognized by OCR engines such as Tesseract 4.0. Each preprocessing method contributes to the improvement of image quality, both in terms of visual readability and clarity of text contours [26].

Grayscale conversion and brightness adjustment are important initial steps in OCR because color information is not really needed for text recognition. By converting the image to grayscale, focus is put on light intensity and contrast, which helps to distinguish the text from the background. After conversion, brightness adjustments are made to reduce the overexposure effect that often occurs, especially if the image is taken under strong or uneven lighting. Grayscale images are digital images that have only one channel value in each pixel, in other words, red = green = blue

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

part values [27]. Too high brightness can cause text to look blurry or unreadable, so this adjustment ensures the text remains clear.

Highlight detection and the application of Gaussian blur are important steps to address overexposed areas that may interfere with the readability of the text. By detecting overexposed areas, we can make further adjustments, where Gaussian blur is used to soften the light intensity in those areas without harming the clarity of the text. This technique helps to maintain the sharpness of the text while reducing distractions from unwanted lighting, thus improving the overall image quality.

CLAHE (Contrast Limited Adaptive Histogram Equalization) is a contrast enhancement technique that is very effective in maximizing text readability in areas of low contrast [28], [29]. This CLAHE process helps to darken the boundaries of nutrient labels to black or near-black values along their contours, which makes a clear difference on lighter backgrounds . This technique is particularly useful when dealing with images that have inconsistent contrast variations, such as in images of nutrient tables with small text over dark or light backgrounds

Denoising, or noise removal, is another important step in preprocessing. Noise can be caused by many factors, such as poor image quality or random patterns in the image. The presence of noise can make text difficult to read and cause errors in the OCR process. By using methods such as Non-Local Means Denoising, noise in images can be minimized without losing important text details. This technique is effective in maintaining the clarity of the text structure while removing the surrounding distracting elements [30].

Otsu's thresholding is a preprocessing method used to convert grayscale images to binary (black-and-white) by automatically determining the best threshold value. This technique maximizes the variance between the two classes of pixels (light and dark), which helps to separate the text from the background. After this process, an inversion check is performed to ensure that the text is black on a white background, which is the optimal format for OCR. A basic understanding of thresholding states that the left histogram represents an image f (x, y), which consists of light objects on a dark background [27]. Mathematically, it can be described in the following formula:

$$\sigma_{\omega}^2(t) = \omega 0(t)\sigma_0^2(t) + \omega 1(t)\sigma_1^2(t) \quad (1)$$

Where t is the threshold, σ is the within-class variance and ω is the share of pixels belonging to each class [31].

The last step is to convert the image back to RGB format so that it can be processed by Tesseract OCR which requires images in this format. Once all these preprocessing steps are completed, the nutrition table image has the ideal visual quality for text recognition by OCR which greatly improves the accuracy of text extraction from images.

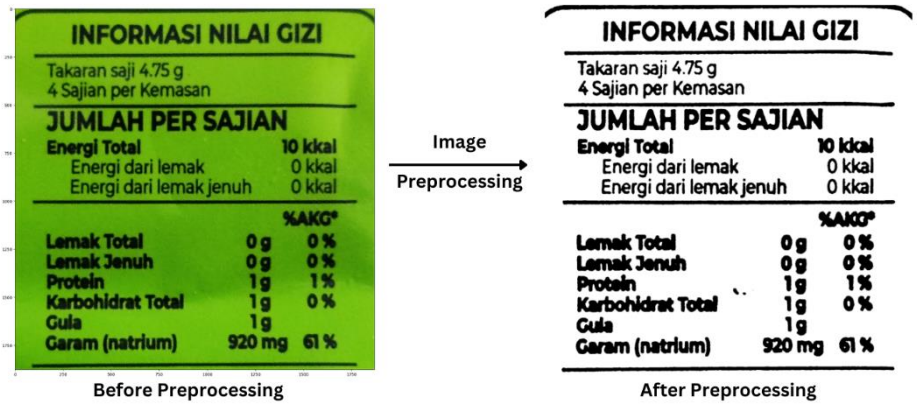


Figure 3. Final Preprocessing Result

Figure 3 shows an example of the different photo conditions before and after preprocessing. Before preprocessing, the nutrition table image has a purple background with white text, which makes text recognition more difficult due to low contrast and uneven lighting. After preprocessing, the image is converted to black-and-white with enhanced contrast, so that the text can be seen more clearly and separated from the background. The visual quality of the image is significantly improved through this process, which makes the image more readable by the OCR engine and can increase the accuracy of text extraction.

2.4 Tesseract OCR 4.0

Optical Character Recognition (OCR) translates written text images into editable text machines [25]. Tesseract as an OCR engine is one of the most widely used open-source engines [12]. Tesseract OCR 4.0 is an open-source optical character recognition (OCR) engine developed by Google, which uses a deep learning model based on Long Short-Term Memory (LSTM). This version brings significant improvements over previous versions with the ability to recognize more complex text, including handwriting, stylized fonts, and text on low-quality images. The use of LSTM architecture allows Tesseract 4.0 to more accurately recognize character sequences in texts of various formats. In addition, Tesseract 4.0 supports more than 100 languages, including languages with right-to-left layouts, which increases the flexibility of its use in various scenarios. Although it takes longer to process than previous versions, the

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

higher accuracy of text recognition results, especially in documents with a lot of noise, makes Tesseract 4.0 a reliable choice for many modern OCR needs [30].

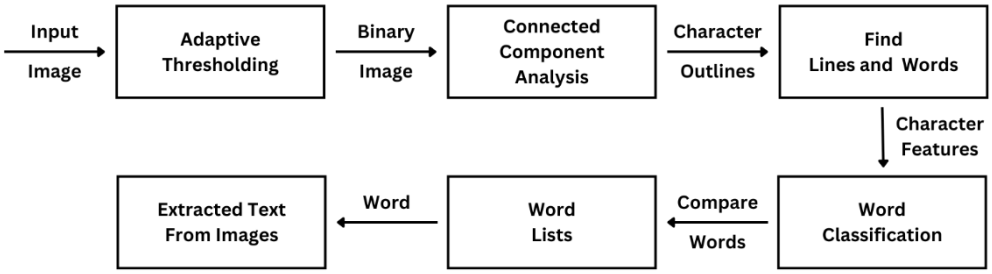


Figure 4. Basic How Tesseract OCR Works

As depicted in figure 5, initially, adaptive thresholding is used to read the input image and convert it into a binary image, or grayscale image. Character outlines are generated through connected component analysis on this binary image. This helps to distinguish the outline from the word in the image. The next challenge is to distinguish each letter present in the word. After that, each character is fed into the character recognition classifier model [31], [32].

2.5 Postprocessing Data

Post-processing is an important stage in improving the quality of text extraction results from OCR (Optical Character Recognition). Although OCR technology has undergone many advancements, the results often still contain various errors, especially when dealing with misrecognized characters, formatting errors, or errors in numeric units. Therefore, the application of post-processing techniques is necessary to correct and clean the extracted text, so that it is more accurate and ready for further analysis [33].

One of the main approaches in post-processing is rule-based correction, where certain rules are applied to correct common errors that appear in OCR results. For example, errors in units, such as writing "mq" where it should be "mg" or "cal" where it should be "kcal", can be automatically corrected using a dictionary containing a list of common errors and their corresponding corrections. This technique can also be used to correct character recognition errors, such as the number "0" which is often detected as the letter "O", or "1" which is detected as "I" or "l" [34].

Another method often used in post-processing is contextual analysis, where misrecognized words or numbers are corrected based on the context of the sentence or data pattern. For example, numbers found near certain units such as "g" or "mg" most likely refer to weight values, so errors in those units can be identified and corrected. This context-based analysis is especially important in nutritional data, where every small error in numbers and units can lead to misinterpretation.

2.6 Word and Character Error Rate

The final stage where after post-processing the image, an error rate calculation will be performed to determine the effectiveness of OCR during image processing. The author uses Word Error Rate (WER) and Character Error Rate (CER) as the calculation of reading errors in the resulting OCR process. All the formula calculations are calculated using the following formula, where for WER is:

$$WER = \frac{Levenshtein\ Distance\ (predicted\_words, ground\_truth\_words)}{Number\ of\ words\ in\ ground\ truth} \quad (2)$$

Meanwhile, CER has a similar formula where if WER is calculated based on existing words, CER is calculated based on each character. The formula of CER itself is as follows:

$$CER = \frac{Levenshtein\ Distance\ (predicted\_charcaters, ground\_truth\_characters)}{Number\ of\ characters\ in\ ground\ truth} \quad (3)$$

In both formulas, the Levenshtein Distance counts the number of edits (insertions, deletions, or substitutions) needed to transform the predicted text into the ground truth, while the number of words or characters in the ground truth normalizes the error rate. A lower WER and CER indicates higher accuracy in the OCR output, reflecting fewer discrepancies compared to the actual text.

3. RESULT AND DISCUSSION

3.1 Research Workflow

It is important to understand that this process consists of several stages that support each other to ensure accuracy in the extraction of nutrition information. Each stage, from object detection to text result correction, is designed to

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

maximize the clarity and accuracy of the data extracted from nutrition labels. The following is an explanation of each stage shown in the research flow figure below.

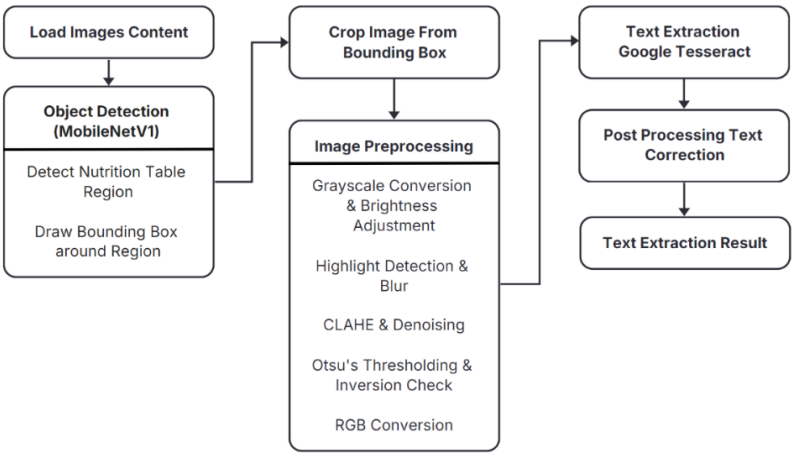


Figure 5. Text Extraction Process Flowchart

Figure 5 depicts the research method flowchart of the text extraction process from nutrition table images, starting with table area detection using the MobileNetV1 model to identify and bound the relevant area. Once the area is detected, the image is cropped based on the bounding box and then goes through a preprocessing stage that includes conversion to grayscale, brightness adjustment, highlight detection, blurring, CLAHE application and denoising, and Otsu thresholding with inversion checking before being converted back to RGB format. The preprocessing results are then extracted using Google Tesseract 4.0 to extract the text, which is then refined through postprocessing to correct extraction errors and produce a more accurate final text.

3.2 Object Detection with Bounding Box

Object detection performed on the food nutrition label packaging will be marked with a green bounding box, this aims to make text reading with OCR in the future can be focused only on the food nutrition label, so that the accuracy of text reading becomes clearer without any interference from other image side texts. Another purpose of this bounding box is for cropping the image at a later stage. This method applies the architecture model of MobileNetV1 as a layer for nutrition label detection on processed images.

Figure 6. Nutrition Label Bounding Box Detection Result

Figure 6 represents that all detected images that have a nutrition label will be detected with a bounding box. When inputting an image that does not have a nutrition table, the bounding box will not appear due to the accuracy factor below the predetermined threshold.

3.3 Cropping Image

The data that already has the coordinates (ymin, xmin, ymax, xmax) of the bounding box will be processed as the cropping needs of a certain part of the image based on the object detection. This function works by accepting an image as input, along with a set of bounding boxes indicating the detected area, and a score indicating the model's accuracy of the detection.

a. The first process is to take the image dimensions (height, width) to calculate the actual coordinates of the bounding box containing the detected object.
b. Then, for each bounding box, it will be checked whether its object detection score is greater than the predefined threshold (threshold 0.5), and the process will continue by calculating the actual position of the bounding box in the image, converting the relative coordinates (between 0 and 1) into real pixels.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

c. After calculating the top, bottom, left, and right coordinates of the bounding box, the part of the image inside these areas is cropped and returned as a cropped image.
d. If there are no eligible detections (below the score threshold), the result will return an empty value or None, indicating that no image was cropped.

In this way, the function ensures that only the relevant area, i.e. the detected nutrition label, is cut out of the image, thus facilitating subsequent processes such as text reading with OCR.

Figure 7. Cropping Image Result of Bounding Box

Figure 7 explains the cropping result on the bounding box threshold that has been determined based on the coordinates obtained from object detection on the food nutrition label packaging.

3.4 Preprocess Image

This stage has an important role before text extraction using OCR. The OCR preprocessing process on food nutrition labels starts with converting the image to grayscale format, which simplifies the image to light intensity only. The brightness is then adjusted using the values α (alpha) = 0.6 for contrast and β (beta) = 50 for brightness, to reduce the effect of overexposure. After that, areas with very bright intensity (240-255) are detected to identify highlights that may interfere with text reading, and Gaussian blur with a (5, 5) kernel is applied to soften these areas. Then contrast is enhanced through CLAHE (Contrast Limited Adaptive Histogram Equalization) method followed by denoising to remove noise. After that, Otsu's Thresholding is applied to convert the image to binary, so that the text is more easily recognized from the background. Finally, the binary image is converted back to RGB format for OCR preparation with Tesseract, ensuring the resulting image is more optimized and easily readable by the OCR system.

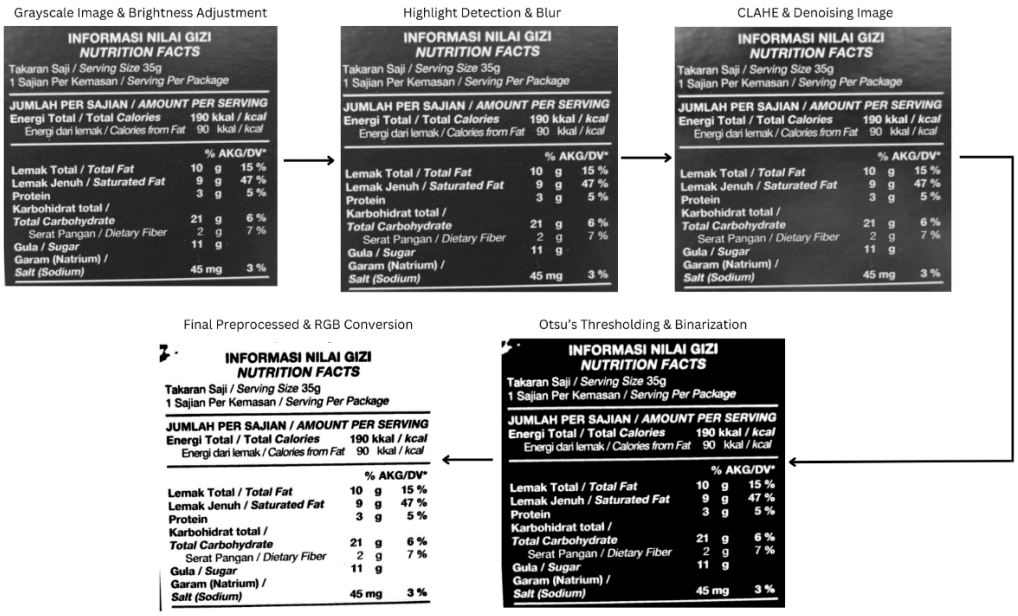


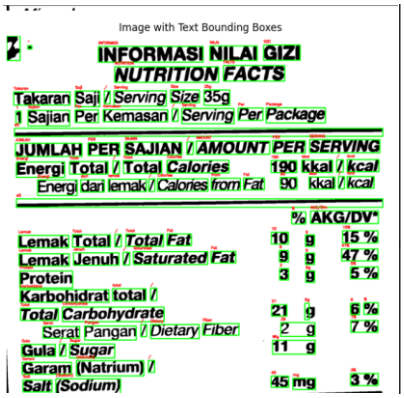
Figure 8. Result of Image Preprocessing Application

Figure 8 explains the stages performed in image preprocessing. There are 5 stages of the sequence and each method has a different process. Starting from Grayscale Image and Brightness Adjustment, Highlight Detection and Blur, CLAHE and Image Denoising, Otsu's Thresholding and Binarization, and finally the final result of preprocess and returned to RGB values for the image.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

3.5 Text Extraction

After preprocessing the image, the extraction stage was performed on the image using Tesseract's OCR engine. The author uses the psm6 segmentation model to detect a block of text in one line in the processed image. Each text detected by OCR will continue to be iterated and drawn its bounding box until the engine detects that no more text is generated at the end.




```
, " INFORMASI NILA! GIZI
NUTRITION FACTS
Takaran Saji / Serving Size 35g
1 Sajian Per Kemasan / Serving Per Package
eS
JUMLAH PER SAJIAN / AMOUNT PER SERVING
Energi Total / Total Calories 190 kkal / kcal
Energi dari lemak / Calories from Fat 90 kkal / kcal
eS
% AKG/DV*
Lemak Total / Total Fat 10 g 15%
Lemak Jenuh / Saturated Fat 9 g 47%
Protein 3 9g 5%
Karbohidrat total /
Total Carbohydrate 21 9g 6 %
Serat Pangan / Dietary Fiber 29 7%
Gula / Sugar Wig
Garam (Natrium) /
Salt (Sodium) 45 mg 3%
```

Figure 9. Text Extraction Results after Preprocessing

Figure 9 is a visualization of the results of text extraction from images by OCR, all extraction results will be stored in the form of text files. From the results of the word extraction formed, postprocessing will then be carried out on the image to evaluate word errors that often appear in the nutritional label case study tested.

3.6 Postprocess Image

This postprocess stage aims to improve the text generated from OCR (Optical Character Recognition) by applying various correction rules to correct common errors in the generated text. This process uses regular expressions (regex) and correction dictionaries to modify the text to make it more accurate and in accordance with the expected format, especially for text containing nutritional information or units.

Before Post Process

```
, " INFORMASI NILA! GIZI
NUTRITION FACTS
Takaran Saji / Serving Size 35g
1 Sajian Per Kemasan / Serving Per Package
eS
JUMLAH PER SAJIAN / AMOUNT PER SERVING
Energi Total / Total Calories 190 kkal / kcal
Energi dari lemak / Calories from Fat 90 kkal / kcal
eS
% AKG/DV*
Lemak Total / Total Fat 10 g 15%
Lemak Jenuh / Saturated Fat 9 g 47%
Protein 3 9g 5%
Karbohidrat total /
Total Carbohydrate 21 9g 6 %
Serat Pangan / Dietary Fiber 29 7%
Gula / Sugar Wig
Garam (Natrium) /
Salt (Sodium) 45 mg 3%
```

After Post Process

```
INFORMASI NILA GIZI
NUTRITION FACTS
Takaran Saji / Serving Size 35g
1 Sajian Per Kemasan / Serving Per Package
eS
JUMLAH PER SAJIAN / AMOUNT PER SERVING
Energi Total / Total Calories 190 kkal / kcal
Energi dari lemak / Calories from Fat 90 kkal / kcal
eS
% AKG/DV
Lemak Total / Total Fat 10 g 15%
Lemak Jenuh / Saturated Fat 9 g 47%
Protein 3 9g 5%
Karbohidrat total /
Total Carbohydrate 21 9g 6 %
Serat Pangan / Dietary Fiber 29 7%
Gula / Sugar 11g
Garam Natrium /
Salt Sodium 45 mg 3%
```

Figure 10. Comparison of Text Extraction Results Before and After Post Processing

Figure 10 is the result of text comparison before post process and after image post process. The author uses two dictionaries that are integrated between unit corrections and word corrections. These two corrections are applied simultaneously in one process. After separating the text into words, the function checks each word for two things:

a. If the word is a unit error, then the unit is corrected with the correct format.
b. If the word is found in the word correction dictionary, it is replaced with the correct word.

3.7 Results and Error Rate Calculation

The text extraction process from food nutrition labels utilized Tesseract OCR combined with an image preprocessing pipeline to enhance the text quality. Additionally, bounding box detection was implemented using MobileNetV1 model to localize regions containing nutritional information, ensuring that the OCR focuses specifically on the text areas in nutrition label. After text extraction, a post-processing step was applied to correct common OCR errors, enhancing the alignment of extracted text with the original content. By measuring WER and CER across a sample set of nutrition labels, we aim to assess the effectiveness of these preprocessing and post-processing steps, ultimately reflecting the accuracy and reliability of the OCR process.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

Table 1. WER and CER Result Score

Image	Word Error Rate Score	Character Error Rate Score
sample_1.jpg	0.069767	0.016713
sample_2.jpg	0.000000	0.000000
sample_3.jpg	0.047619	0.005540
sample_4.jpg	0.047619	0.036312
sample_5.jpg	0.023255	0.002785
sample_6.jpg	0.100000	0.013966
sample_7.jpg	0.000000	0.000000
sample_8.jpg	0.097560	0.016620
sample_9.jpg	0.073170	0.008379
sample_10.jpg	0.047619	0.019553

Table 1 describes the results of the WER and CER tests for the entire text extraction process. The author tests using 10 different sample images of nutrition labels, each of which has a score of WER and CER. It can be seen in the table above for the maximum value of the error threshold limit on WER is 10% and on CER 1.6%. The lower the percentage of WER and CER produced means that the extraction value produced by OCR is more accurate.

4. CONCLUSION

The conclusion of this study confirms that the implementation of OCR with the Tesseract engine, combined with bounding box detection using MobileNetV1, is effective in extracting nutritional information from nutrition labels on food packaging. With low error rates, such as an average Word Error Rate (WER) of 0.0506609 and an average Character Error Rate (CER) of 0.0119868 from 10 tested samples, the system shows great potential in helping consumers obtain nutritional information automatically. The process performed prior to text extraction, namely the preprocessing stage, includes several important steps to ensure the image is ready to be processed by the OCR engine. These include converting the image to grayscale to focus on light intensity, adjusting the brightness to make the text more clearly visible, applying CLAHE to increase contrast in areas with low contrast backgrounds, and noise reduction to remove visual distractions that could affect text recognition accuracy. After the text is successfully extracted using Tesseract OCR, a postprocessing stage is applied to correct any errors that may have occurred during extraction. This stage uses rule-based methods and contextual analysis to correct common errors such as misrecognition of characters or unit formats. Thus, the final result becomes more accurate and ready to be used in various applications. The system is expected to improve consumers' accessibility and understanding of nutrition information, which will ultimately contribute to increased public health awareness, especially in terms of making better decisions regarding food choices.

ACKNOWLEDGMENT

We would like to express our deepest gratitude to God Almighty for His endless mercy and grace, enabling us to complete this research successfully. Our sincere thanks also go Universitas Dian Nuswantoro for their invaluable support throughout the research process. We are especially grateful to our mentors, colleagues, and friends for their guidance, expertise, and encouragement, which have been instrumental in shaping this work. We hope that the findings of this research will contribute to the advancement of science and technology, bringing meaningful benefits to the wider community and inspiring future innovations in this field.

REFERENCES

[1] A. Mahfudhin and P. Kurnia, “Hubungan pengetahuan dengan perilaku membaca label informasi nilai gizi pada ahli gizi di Surakarta,” *Ilmu Gizi Indonesia*, vol. 5, no. 1, p. 47, Aug. 2021, doi: 10.35842/ilgi.v5i1.209.

[2] Q. A. Huda and D. R. Andrias, “SIKAP DAN PERILAKU MEMBACA INFORMASI GIZI PADA LABEL PANGAN SERTA PEMILIHAN PANGAN KEMASAN,” *Media Gizi Indonesia*, vol. 11, no. 2, p. 175, Jan. 2018, doi: 10.20473/mgi.v11i2.175-181.

[3] Y. D. Sari and R. Rachmawati, “KONTRIBUSI ZAT GIZI MAKANAN JAJANAN TERHADAP ASUPAN ENERGI SEHARI DI INDONESIA (ANALISIS DATA SURVEY KONSUMSI MAKANAN INDIVIDU 2014) [FOOD AWAY FROM HOME (FAFH) CONTRIBUTION OF NUTRITION TO DAILY TOTAL ENERGY INTAKE IN INDONESIA],” *Penelitian Gizi dan Makanan (The Journal of Nutrition and Food Research)*, vol. 43, no. 1, pp. 29–40, Sep. 2020, doi: 10.22435/pgm.v43i1.2891.

[4] K. P. A. Nugroho, R. R. M. D. Kurniasari, and T. Noviani, “GAMBARAN POLA MAKAN SEBAGAI PENYEBAB KEJADIAN PENYAKIT TIDAK MENULAR (DIABETES MELLITUS, OBESITAS, DAN HIPERTENSI) DI WILAYAH KERJA PUSKESMAS CEBONGAN, KOTA SALATIGA,” *Jurnal Kesehatan Kusuma Husada*, pp. 15–23, Jan. 2019, doi: 10.34035/jk.v10i1.324.

[5] T. Seviana, *Profil Kesehatan Indonesia 2023*. Kementrian Kesehatan Republik Indonesia, 2023.

[6] M. J. Christoph and R. An, “Effect of nutrition labels on dietary quality among college students: a systematic review and meta-analysis,” *Nutr Rev*, vol. 76, no. 3, pp. 187–203, Mar. 2018, doi: 10.1093/nutrit/nux069.

Building of Informatics, Technology and Science (BITS)
Volume 6, No 3, Desember 2024 Page: 1403−1412
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i3.6107

[7] N. S. Pratt, B. D. Ellison, A. S. Benjamin, and M. T. Nakamura, "Improvements in recall and food choices using a graphical method to deliver information of select nutrients," *Nutrition Research*, vol. 36, no. 1, pp. 44–56, Jan. 2016, doi: 10.1016/j.nutres.2015.10.009.
[8] Y. Shah, "Delving Deep into NutriScan: Automated Nutrition Table Extraction and Ingredient Recognition," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 11, pp. 1596–1601, Nov. 2023, doi: 10.22214/ijraset.2023.56852.
[9] A. Kaderabek, "Exploring Optical Character Recognition (OCR) as a Method of Capturing Data from Food-Purchase Receipts," *Surv Methods Insights Field*, vol. 1, no. 3, pp. 1–15, Nov. 2023.
[10] D. Sporici, E. Cușnir, and C.-A. Boiangiu, "Improving the Accuracy of Tesseract 4.0 OCR Engine Using Convolution-Based Preprocessing," *Symmetry (Basel)*, vol. 12, no. 5, p. 715, May 2020, doi: 10.3390/sym12050715.
[11] A. Inbasekaran, R. K. Gnanasekaran, and R. Marciano, "Using Transfer Learning to contextually Optimize Optical Character Recognition (OCR) output and perform new Feature Extraction on a digitized cultural and historical dataset," in *2021 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2021, pp. 2224–2230. doi: 10.1109/BigData52589.2021.9671586.
[12] H. Seitaj and V. Elangovan, "Information Extraction from Product Labels: A Machine Vision Approach," *International Journal of Artificial Intelligence & Applications*, vol. 15, no. 2, pp. 57–76, Mar. 2024, doi: 10.5121/ijaia.2024.15204.
[13] N. A. Putri Kamis and O.-K. Shin, "OCR-Based Safety Check System of Packaged Food for Food Inconvenience Patients," *Journal of Digital Contents Society*, vol. 21, no. 6, pp. 1025–1032, Jun. 2020, doi: 10.9728/dcs.2020.21.6.1025.
[14] T. Hegghammer, "OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment," *J Comput Soc Sci*, vol. 5, no. 1, pp. 861–882, May 2022, doi: 10.1007/s42001-021-00149-1.
[15] P. Jain, Dr. K. Taneja, and Dr. H. Taneja, "Which OCR toolset is good and why? A comparative study," *Kuwait Journal of Science*, vol. 48, no. 2, Apr. 2021, doi: 10.48129/kjs.v48i2.9589.
[16] M. Tomaschek, "Evaluation of off-the-shelf OCR technologies," Universitass Masarykiana, 2017.
[17] T. Palwankar and K. Kothari, "Real Time Object Detection using SSD and MobileNet," *Int J Res Appl Sci Eng Technol*, vol. 10, no. 3, pp. 831–834, Mar. 2022, doi: 10.22214/ijraset.2022.40755.
[18] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017.
[19] H. Wang, H. Qian, S. Feng, and W. Wang, "L-SSD: lightweight SSD target detection based on depth-separable convolution," *J Real Time Image Process*, vol. 21, no. 2, p. 33, Apr. 2024, doi: 10.1007/s11554-024-01413-z.
[20] B. T. Naik and M. F. Hashmi, "MobileNet + SSD: Lightweight Network for Real-Time Detection of Basketball Player," 2023, pp. 11–19. doi: 10.1007/978-981-19-8742-7_2.
[21] V.-S. Vu and D.-N. Nguyen, "Application of MobileNet-SSD Deep Neural Network for Real-Time Object Detection and Lane Tracking on an Autonomous Vehicle," 2022, pp. 559–565. doi: 10.1007/978-3-030-91892-7_53.
[22] E. Suharto, Suhartono, A. P. Widodo, and E. A. Sarwoko, "The use of mobilenet v1 for identifying various types of freshwater fish," *J Phys Conf Ser*, vol. 1524, no. 1, p. 012105, Apr. 2020, doi: 10.1088/1742-6596/1524/1/012105.
[23] N. Gupta and N. M. Khan, "Efficient and Scalable Object Localization in 3D on Mobile Device," *J Imaging*, vol. 8, no. 7, p. 188, Jul. 2022, doi: 10.3390/jimaging8070188.
[24] E. Z. Orji, A. Haydar, İ. Erşan, and O. O. Mwambe, "Advancing OCR Accuracy in Image-to-LaTeX Conversion—A Critical and Creative Exploration," *Applied Sciences*, vol. 13, no. 22, p. 12503, Nov. 2023, doi: 10.3390/app132212503.
[25] D. N. Kholifah, H. M. Nawawi, and I. J. Thira, "IMAGE BACKGROUND PROCESSING FOR COMPARING ACCURACY VALUES OF OCR PERFORMANCE," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 33–38, Mar. 2020, doi: 10.33480/pilar.v16i1.1076.
[26] J. Jayanthi and P. U. Maheswari, "Comparative study: enhancing legibility of ancient Indian script images from diverse stone background structures using 34 different pre-processing methods," *Herit Sci*, vol. 12, no. 1, p. 63, Feb. 2024, doi: 10.1186/s40494-024-01169-6.
[27] T. Wang, G. T. Kim, M. Kim, and J. Jang, "Contrast Enhancement-Based Preprocessing Process to Improve Deep Learning Object Task Performance and Results," *Applied Sciences*, vol. 13, no. 19, p. 10760, Sep. 2023, doi: 10.3390/app131910760.
[28] B. LIU and J. LIU, "Overview of image noise reduction based on non-local mean algorithm," *MATEC Web of Conferences*, vol. 232, p. 03029, Nov. 2018, doi: 10.1051/matecconf/201823203029.
[29] J. Reibring, "Photo OCR for Nutrition Labels Combining Machine Learning and General Image Processing for Text Detection of American Nutrition Labels," 2017.
[30] GitHub, "Tesseract documentation," github.io. Accessed: Oct. 20, 2024. [Online]. Available: https://tesseract-ocr.github.io/
[31] N. Chigali, S. R. Bobba, K. Suvarna Vani, and S. Rajeswari, "OCR Assisted Translator," in *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, IEEE, Jul. 2020, pp. 1–4. doi: 10.1109/ICSSS49621.2020.9202034.
[32] M. Heidarysafa, J. Reed, K. Kowsari, A. C. R. Leviton, J. I. Warren, and D. E. Brown, "From Videos to URLs: A Multi-Browser Guide To Extract User's Behavior with Optical Character Recognition," Nov. 2018.
[33] S. Drobac and K. Lindén, "Optical character recognition with neural networks and post-correction with finite state methods," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 23, no. 4, pp. 279–295, Dec. 2020, doi: 10.1007/s10032-020-00359-9.
[34] L. L. de Oliveira *et al.*, "Evaluating and mitigating the impact of OCR errors on information retrieval," *International Journal on Digital Libraries*, vol. 24, no. 1, pp. 45–62, Mar. 2023, doi: 10.1007/s00799-023-00345-6.