Cek Plagiasi Windra SIET-2017

by Windra .

Submission date: 29-Jan-2019 10:00AM (UTC+0700)
Submission ID: 1069885204
File name: SIET-2017.pdf (692.13K)
Word count: 2586
Character count: 13398

2017 International Conference on Sustainable Information Engineering and Technology (SIET)

Android Based Application for Recognition of Indonesian Restaurant Menus Using Convolution Neural Network

Windra Swastika
Faculty of Science and Technology
Universitas Ma Chung
Malang, Indonesia
windra.swastika@machung.ac.id

Yoko
Faculty of Science and Technology
Universitas Ma Chung
Malang, Indonesia

Hendry Setiawan
Faculty of Science and Technology
Universitas Ma Chung
Malang, Indonesia

Moch. Subianto
Faculty of Science and Technology
Universitas Ma Chung
Malang, Indonesia

***Abstract*— Indonesia has a lot of typical food which is rich in flavor. However, compared with countries such as China, Thailand or India, Indonesian food is not widely known by the international community. One of the factors that cause Indonesian food is less popular is that the food information cannot be accessed easily by foreign travelers during a visit to Indonesia. Food name is written on the restaurant menu can be read, but not really showing the information about the food, such as ingredients, how the food is prepared or a picture of the food itself. Hence, it is required to facilitate foreign travelers who want to find out information about Indonesian food based on restaurant menus. An Android-based application has been developed to show food information. The application captures a text on the restaurant menus, process and recognize the text using convolution neural network (CNN). The recognized text is then matched with predetermined database to show the information about the food. The application was able to recognize 100% of the menus when the menus use Sans Serif Font. However, the accuracy dropped into 56% when the menus use Times New Roman.**

Keywords*— *menu recognition; optical character recognition; convolution neural network

I. Introduction

Indonesia has a lot of typical food which is rich in flavor. However, compared with countries like China, Thailand, India, and Korea, Indonesian food is not widely known by the international community. The number of restaurants that serve Indonesian food in America for example, less than 20 outlets, while a restaurant serving Chinese food in the United States more than 30,000 outlets.

One of the factors that cause Indonesian food is less popular is that the food information cannot be accessed easily by foreign travelers during a visit to Indonesia. Food name is written on the menu can be read, but not really showing the information about the food, such as ingredients, how the food is prepared or a picture about the food itself. Hence, it is required to facilitate foreign travelers who want to find out information about Indonesian food based on restaurant menus. Smartphones equipped with a camera can be used as a tool to provide information about Indonesian food. An Android-based application will be developed to show food information. The application captures a text on the restaurant menus and processes it. Using the results of the image processing, character recognition technology or Optical Character Recognition (OCR) can be used to recognize text from an image of a restaurant menu. The text that successfully recognized by OCR is matched to a database and send back to the user to show detail information about the food.

In this study, we developed a smartphone application to recognize Indonesian restaurant menus. We choose Android as the platform since it is the most widely used worldwide mobile operating system. The number of Android smartphone users reached 87 million in 2016 and still continues to grow. For image recognition, we used Convolution Neural Network (CNN).

II. Methods

In this study, we use raw image captured from a smartphone camera as an input. The image is then processed using grayscaling, noise removal dan image binarization to produce noiseless binary image. The binary image is used as an input for character recognition to obtain a word or text query for database. Before we implement our method, we design the system using use case diagram, activity diagram, and design user interface. Finally, we test and validate our application using black box method. Figure 1 shows the diagram of our research methodology.

978-1-5386-2182-0/17/$31.00 ©2017 IEEE

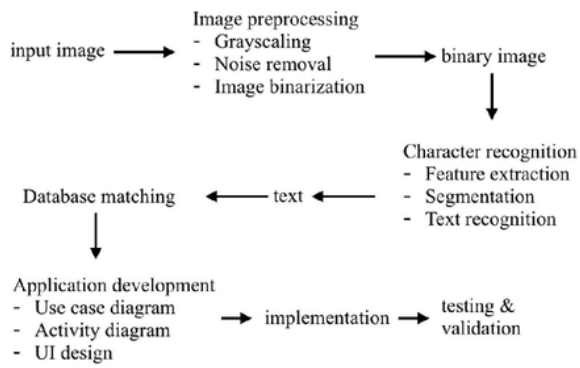


Fig. 1. Research methodology

There are three main processes to recognize the text from the restaurant menu, namely: image preprocessing, image recognition and database matching.

A. Image Preprocessing

Input image for the system is a raw RGB image. It is required to be preprocessed before fed into character recognition. The preprocessing steps are: (1) Grayscaling, (2) Noise removal and (3) Image binarization. Grayscaling is the initial step after an image captured by the smartphone camera. Converting a color image into a gray image with decolorize algorithm method is quite effective in maintaining the contrast of the image [1]. Decolorize method algorithm receives input RGB image (Ri, Gi, Bi) ∈ [0, 1] 3 and change into Ti ∈ [0, 1]. The value of R, G and B of each pixel is converted into color values (Y, P, Q) by using (1):

$$\begin{bmatrix} Y_i \\ P_i \\ Q_i \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.5870 & 0.1140 \\ 0.5000 & 0.5000 & -1.000 \\ 1.000 & -1.000 & 0.000 \end{bmatrix} \begin{bmatrix} R_i \\ G_i \\ B_i \end{bmatrix} \quad (1)$$

YPQ color channels produces a grayscale image Yi ∈ (ymin, ymax) = [0, 1]. After grayscaling, next operation is noise removal using smoothing. By using a smoothing operation, a grayscale image will be blurred to remove detail and noise. The method used in this smoothing process is Gaussian blur that has a smoothing kernel as shown in [2]. In 2D plane, Gaussian distribution has the formula as shown in 2.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

Where x and y are are the horizontal and vertical distance from the point of origin and σ is the standard deviation of the Gaussian distribution. The value of the Gaussian distribution is used as a convolution matrix that is applied to the image. Estimated discrete convolution for a 3x3 matrix with σ = 1 are:

$$\begin{bmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{bmatrix} \quad (3)$$

Final step of image preprocessing is image binarization. The purpose of image binarization is to convert grayscale image into binary image (black and white image). In this study, we use adaptive threshold method to convert grayscale image from previous step into binary image. Unlike global threshold, adaptive threshold uses value T as a threshold value that depends on the neighbor pixels.

B. Image recognition

To recognize the preprocessed image, we use CNN. CNN is known to have better accuracy compared to artificial neural network (ANN) (3, 4, 5). Figure 2 shows the architecture of CNN to be applied for image recognition process.

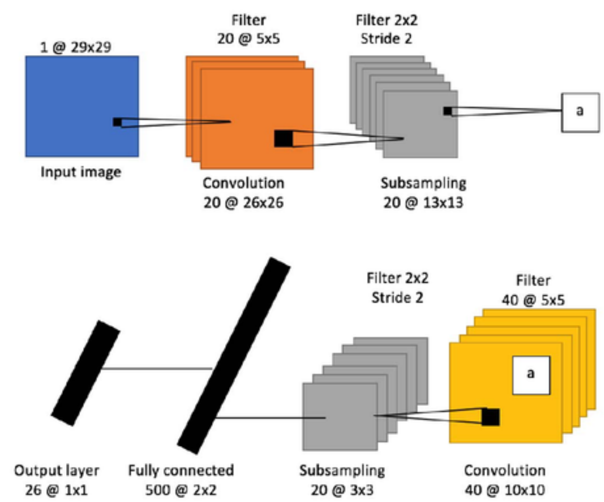


Fig. 2. CNN architecture for the application

In Figure 2, the initial part of the CNN architecture is the input layer which is a preprocessed image data with the size of 29x29. The next process is a convolution layer that convolves input data using 20 filters of 5x5. The convolution process yields 20 feature maps with size 26x26. Each features map will be subsampled with filter 2x2 and stride 2. This step will reduce the dimensionality of each feature map into 13x13. As shown in Figure 2, the results of subsampled feature maps forwarded into 2nd convolution layer using 40 filters with size 5x5 to produces a total of 40 feature maps. Max pooling operation (subsampling) is also performed on each feature maps to produces 20 subsampled feature maps with size 3x3. Each subsampled feature maps act as an input for fully connected layer which is similar as multilayer perceptron in artificial neural network. Next layer after fully connected layer is the output layer. In this case, the output layer consists of 26 node that represents probability of alphabet A to Z.

The training process is a necessary process to obtain correct weight matrices in the CNN. The steps of training process consist of forward propagation (from input image to output layer) and back-propagation process as follows:

1. Initialization step: define filter size and initialize weights with random values.
2. Forward propagation: take image as input, move forward to 1st convolution, 1st subsampling, 2nd convolution, 2nd subsampling, fully connected layer, output layer.
3. Calculate the total error of output layer compared to desired output.
4. Use backpropagation to adjust the weights.
5. Repeat step 2-4 until stopping criteria is met.

C. Database Matching

SQLite is database management system library provided by Android Operating System. The application uses SQLite to store offline data and information about the food. The information provided in the database are:

1. Name
2. The origin
3. Ingredients
4. Meals information
5. Picture

The name of the food in the database will the main search query after the image successfully translated into a text.

D. Flow of the Application

Figure 3 shows use case diagram of the application.

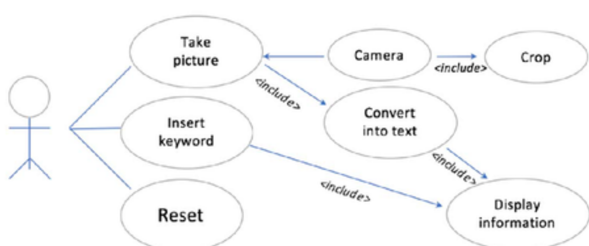


Fig. 3. Use case diagram of the application

There are 3 main menus on the application: Take picture, Insert keyword and Reset. Take Picture will launch the smartphone camera activity to let user take picture of restaurant menu. The obtained picture is manually cropped to focus on a word that user wants to get information about. Once it is cropped, the application will convert into text and match with the database to display information. The user is also able to insert keyword manually to get information. This is the case when the system failed to recognize text obtained from camera.

Figure 4 shows the flow diagram of the application.

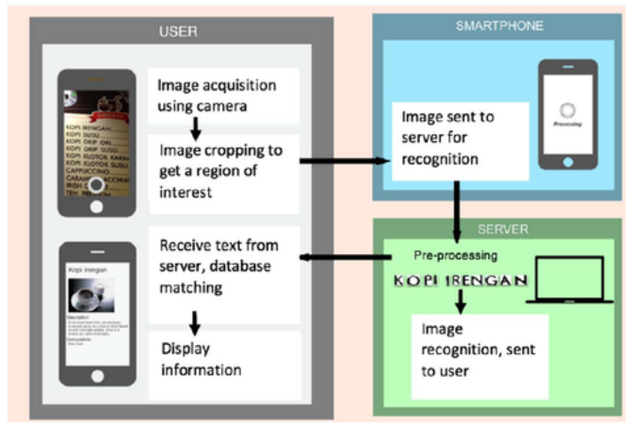
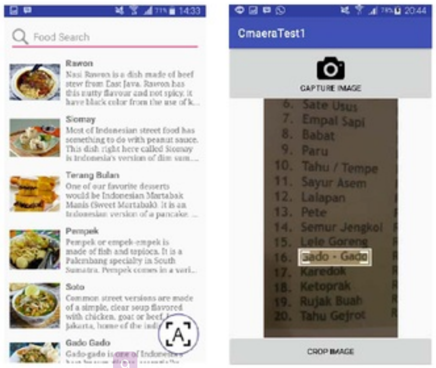


Fig. 4. Diagram flow of application

The application starts by showing a list of menu and a button to take image from camera. User is required to select region of interest by cropping the image. The cropped image is sent to server to perform pre-processing and image recognition using CNN. The result of image recognition is sent back to user in the form of text. The text will be matched with local database to show information about the food.

III. EXPERIMENTAL RESULTS

The user interface of the application consists of 3 main functions as shown in the figure 4. The first function is search function. Search function is used to search food information by entering food name manually (Figure 4a). The second function is to capture image from restaurant menu and manually crop it to obtain region of interest (Figure 4b) and the last function is to show food information based on the manual search or image recognition (Figure 4c).



(a) (b)

(c)

Fig. 5. Interface of the application; (a) manual food search; (b) capture and crop the image; (c) food information

In this application, all image processing and text recognition were done on the server. Figure 5a shows an example of cropped image that sent to the server and preprocesses it using MatLab 2016b which is running continuously on the server. Figure 4b shows the results of preprocessing which consists of grayscaling, noise removal and adaptive threshold. The result of preprocessing is then segmented to obtain the characters. Each character is normalized to matrix with size 29x29 and forwarded to CNN engine to recognize or convert into a text.

(a)

(b)

Fig. 6. (a) Cropped image that received by server; (b) Results of preprocessing

The converted text is sent back to the client's smartphone and matched with local database stored in application. Since the matrix size of character is quite small (29x29), in many cases, characters were misconverted and consequently, the result of final text also mismatch compared to the real name of the food. For example, the word "Gado-Gado" was misconverted into "gadobgado", "Ayam goreng" was misconverted into "ayamgofehs", etc. Therefore, Levensthein distance method [6] is used to measure the similarity between converted text and keyword on the database. The minimum value of Levensthein distance is 0 and the maximum value is 1 which means the converted text has exact same characters and its order with the original food name. In this study, we only show the food information if it has minimum 0.5 similarity value of database matching.

We tested the application using 2 different restaurant menus which consist of more than 20 different food names. The first menu uses Sans Serif font and the second menu uses Times New Roman. All of the food names and information were stored in the SQLite database. Figure 6 shows the restaurant menus we used to test the application.

(a) (b)

Fig. 7. Restaurant menus to test the application

To measure the accuracy, we used Samsung Galaxy Note 4 with 16 MP camera. The camera has autofocus feature and yields an optimal image.

We evaluate 3 kinds of angle to capture the menu, 0º, 30º and 60º. Two lighting conditions, natural light and light bulb were also considered in the experiments. Figure 7 shows an example of 1st menu image taken with 0º angle. The image was cropped (1st row of Figure 7), preprocessed and segmented in several characters (2nd row of Figure 6). The segmented characters will be the input for CNN. The recognition results are shown in the 3rd row of Figure 7. The word "Ketoprak" was recognized as "ketopfdk". Since the word "ketopfdk" is 75% similar as "ketoprak" in the database, the system will show information about "ketoprak".

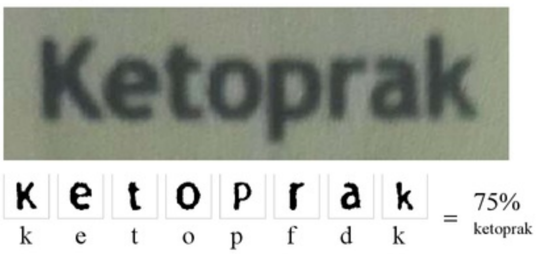


Fig. 8. The word "ketoprak" recognized as "ketopfdk".

Figure 8 shows an example of cropped image from 2nd menu with its recognition accuracy. The font in the second menu is Times New Roman. In this case, the word "Sate Ayam" was recognized as "satcnyam" which has 75% similarity with "sate ayam" in the database.

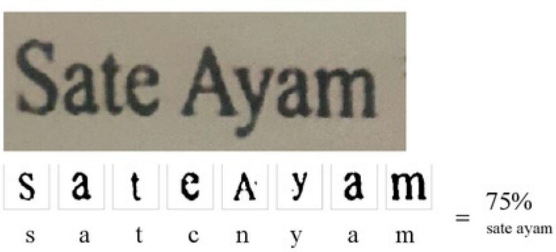


Fig. 9. The word "Sate Ayam" recognized as "satcnyam".

Table 1 shows the accuracy of both restaurant menu with different angles.

TABLE I.

TABLE II. RECOGNITION ACCURACY UNDER DIFFERENT ANGLE AND LIGHT SOURCE

Angle	Light sources	Accuracy (1st menu)	Accuracy (2nd menu)
0º	Natural light	100%	56%
	Light bulb	85%	56%
30º	Natural light	80%	44%
	Light bulb	85%	18%
60º	Natural light	100%	56%
	Light bulb	100%	50%

As shown in table 1, the best accuracy for first menu (sans serif font) is obtained for camera angle of 0º under natural light and 60º camera angel. When using camera angle of 30º, the accuracy dropped to 80% for natural light and 85% for light bulb. In the second menu (times new roman font), the application can only show food information less than 60% of all condition. It seems that times new roman font is the main factor of less accuracy compared to sans-serif font.

IV. CONCLUSION

We have developed an Android-based application for recognition of Indonesian restaurant menus using convolution neural network. The obtain the best accuracy, the menus should take under the natural light condition with camera angle 0º or 60º. Manual cropping also affects the accuracy. Therefore, it is suggested to perform automatic cropping to get consistent and better accuracy.

Although CNN is known to have better accuracy compared to ANN, it is useful to conduct accuracy comparison between CNN and ANN for this application in the future.

REFERENCES

[1] Grundland, Mark, and Neil A. Dodgson. 2005. The decolorize algorithm for contrast enhancing, color to grayscale conversion. University of Cambridge, Technical Report UCAM-CL-TR-649.

[2] Chung, Moo K. 2012. Gaussian kernel smoothing. Lecture notes 1-10.

[3] Zhang, Y., 2015. Deep Convolutional Network for Handwritten Chinese Character Recognition. CS231N Course Project, p. 8.

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[5] Lo, S. C. B., Chan, H. P., Lin, J. S., Li, H., Freedman, M. T., & Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. Neural networks, 8(7), 1201-1214.

[6] Soukoreff, R. W., & MacKenzie, I. S. (2001, March). Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic. In *CHI'01 extended abstracts on Human factors in computing systems* (pp. 319-320). ACM.

Cek Plagiasi Windra SIET-2017

ORIGINALITY REPORT

5%	1%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Bolton Institute of Higher Education Student Paper	1%
2	Samian, , Y.G.Y. Yhuwana, A.H. Zaidan, Supadi, and Sujito. "Application of two-output port fiber coupler as gasoline level sensor", Measurement, 2016. Publication	1%
3	Submitted to Softwarica College of IT & E-Commerce Student Paper	<1%
4	N F Ghazalli, L Yuliati, H O Lintang. "Molecular Self-Assembly of Group 11 Pyrazolate Complexes as Phosphorescent Chemosensors for Detection of Benzene", IOP Conference Series: Materials Science and Engineering, 2018 Publication	<1%
5	ieee.id Internet Source	<1%

6	L.G. Escalante, A.B. Castillo, L.B. Ocana. "Automatic Generation and Publication of Web Services for the Access and Integration of Distributed Data Sources", Sixth Mexican International Conference on Computer Science (ENC'05), 2005 Publication	<1 %
7	en.itu.dk Internet Source	<1 %
8	"Computer Vision – ECCV 2018 Workshops", Springer Nature, 2019 Publication	<1 %
9	Lecture Notes in Computer Science, 2015. Publication	<1 %
10	The Handbooks of Fuzzy Sets Series, 1999. Publication	<1 %
11	Adhiwibawa, Marcelinus A.S., Christian Tantono, Kestrilia R. Prilianti, Monika N.P. Prihastyanti, Leenawaty Limantara, and Tatas.H.P. Brotosudarmo. "Rapid nitrogen determination of soybean leaves using mobile application", 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), 2013. Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Cek Plagiasi Windra SIET-2017

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5