NAME AND RECIPE ESTIMATION OF THAI-DESSERTS BEYOND IMAGE TAGGING

Pakpoom Mookdarsanit[1] and Lawankorn Mookdarsanit[2]
[1]Department of Computer Science, Faculty of Science, Chandrakasem Rajabhat University, 39/1 Ratchadaphisek Road, Huai Khwang, Bangkok 10900, Thailand
[2]Department of Business Computer, Faculty of Management Science, Chandrakasem Rajabhat University, 39/1 Ratchadaphisek Road, Huai Khwang, Bangkok 10900, Thailand

ABSTRACT

Thai-desserts or snacks (Thai: “ขนมไทย”) are included in Thai cuisine. The Food Republic – a well-known organization that gives some information about foods, drinks and snacks, has recommended the 7 Thai-desserts as Khanom Krok (Thai: ขนมครก), Coconut Ice-cream (Thai: ไอศกรีมกะทิสด), Fruity Luk Chub (Thai: ลูกชุบผลไม้), Woon Bai Toey (Thai: วุ้นใบเตย), Tub Tim Krob (Thai: ทับทิมกรอบ), Luem Gluen (Thai: ลืมกลืน) and Bua Loy (Thai: บัวลอย). However, the international outsiders do not dare to have these beautiful and delicious Thai-desserts because of the inconvenience of finding these Thai names and recipes. For example, some outsiders have diabetes and cannot have too much sugar or insulin. This paper combines the “Artificial Intelligence” and “Computer Vision” to build a computer model for tagging the name and recipe of unknown image. Our computer model was done on Convolution Neural Networks (CNN) that provided more than 85% of the accuracy.

KEYWORDS: Dessert Recognition, Image Tagging, Thai Desserts

1. Introduction

The definition of “Thai-dessert(s)” or “snack(s)” is a confectionery course [1] that is always served to the consumers after the main meal in Thai cuisine. The course usually consists of various sweet desserts and their recipes are mostly composed from the country’s natural and agricultural products. The Food Republic [2] – a well-known organization that gives some information about foods, drinks and snacks, has recommended the 7 Thai-desserts as shown in figure 1.

Figure 1 The 7 Thai-desserts recommended by "The Food Republic"

For outsiders (foreigners/ non-Thais), those desserts can be generally tasted in either "luxurious restaurant(s)" or "street food market(s)" in Thailand. Those desserts also can be seen as "Thai-ness (Thai: ความเป็นไทย)" in light of "Thai folk wisdom (Thai: ภูมิปัญญาไทย)" [3]. To put it another way, it is really inconvenient to find those "Thai name(s)" and "recipe(s)" during the travel [4]. The inconvenience of finding those Thai-keywords means the international outsiders do not dare to have those attractive desserts in spite of their beautiful features and delicious tastes. Importantly, some outsiders have so the high level of blood sugar that they have to control the amount of sugar consumption. Considering all factors, this paper proposes a novel "Name and Recipe Estimation of Thai-desserts beyond Image Tagging" that builds a "computer model" for tagging (or estimating) the "name" and "recipe" of unknown image. This paper is the first groundwork that applies the subjects of "Computer Vision" [5, 6] and "Artificial Intelligence" [7, 8] into Thai-desserts. Since these subjects have been successful in previous applications like place recognition [9, 10], facial detection [11, 12], architecture identification [13-15], satellite imageries [16-18], tourism categorization [19] and human detection [20-22]. Likewise, it can be applied in food recognition [23, 24]. The researches about food recognition are a supervised-based computer model that has been gradually increasing since 2017 [25, 26]. As well as the food, it is feasible to build the computer model of desserts. Our model is done by "Convolution (or Convolutional) Neural Networks (CNN)". The "unknown image" is automatically checked to search for its name and recipes by CNN. In conclusion, our computer model provided more than 85% of the accuracy.

The organization of this paper can be divided into 4 sections. The section "CNN-based computer model of name and recipe-tagged images" and "name and recipe estimation of dessert images" are in section 2 and 3, respectively. And the section 4 describes "conclusions".

2. CNN-based Computer Model of Name and Recipe-tagged Images

This section deeply describes our architecture of computer model "Convolutional Neural Networks (CNN)" [27] that is a type of feed-forwarding Neural Networks. Due to the very large number of pixels within a recipe-tagged image, the traditional Neural Networks (also called "Multi Layer Perceptron: MLP") applied in computer vision easily produces very high number of neurons. For example, the "128x128 colored image" has 3 dimensions (in red, green and blue) that would have "128x128x3=49,152 weights and biases". Moreover, the number of input images is defined by "number of features". As opposed to MLP [28], CNN provides the solution by reducing the number of useless pixels. Our architecture of CNN based computer model can be organized into 4 sub-sections: "Input Image", "Convolutions (CONV)", "Max Pooling Layers (MPL)" and "Rectified Linear Units (RELU) with Output Classes", as shown in figure 2.

2.1 Input Image

The name-and-recipe-tagged images are trained to CNN-based computer model. For more convenience, all sizes of the color images are converted or cropped into 128x128x3 pixels. In our experiment, we used 480 recipe-tagged images to build the computer model.

2.2 Convolutions

The Convolutional operation of CNN is used to extract the features from an image before passing to the next "Max Pooling Layers (MPL)". Convolution can improve the free parameters using the "concept of neighbour pixels" [29, 30]. The examples can be visualized in figure 3.

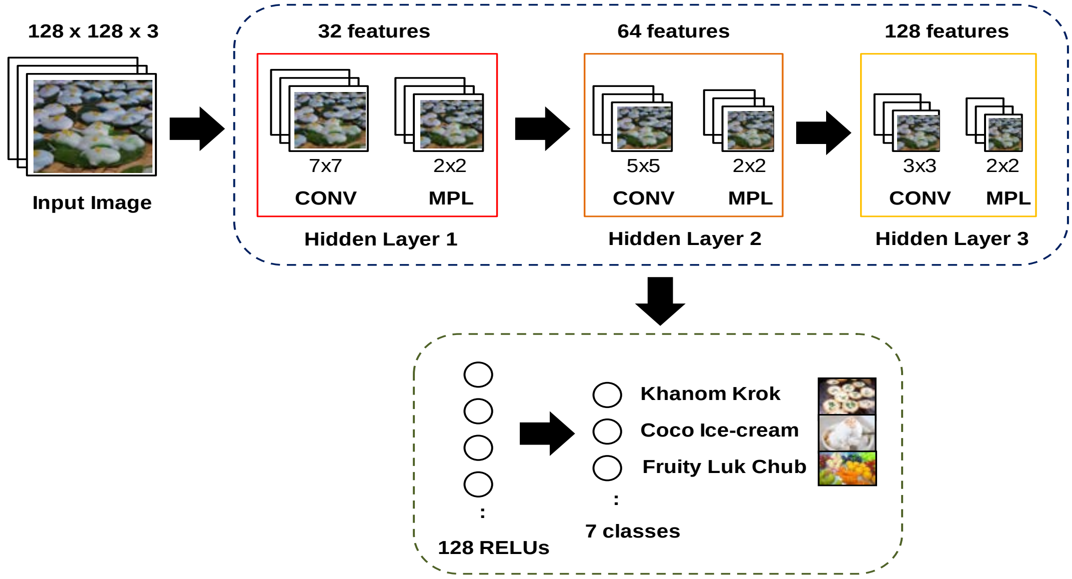


Figure 2 Architecture of CNN-based Computer Model

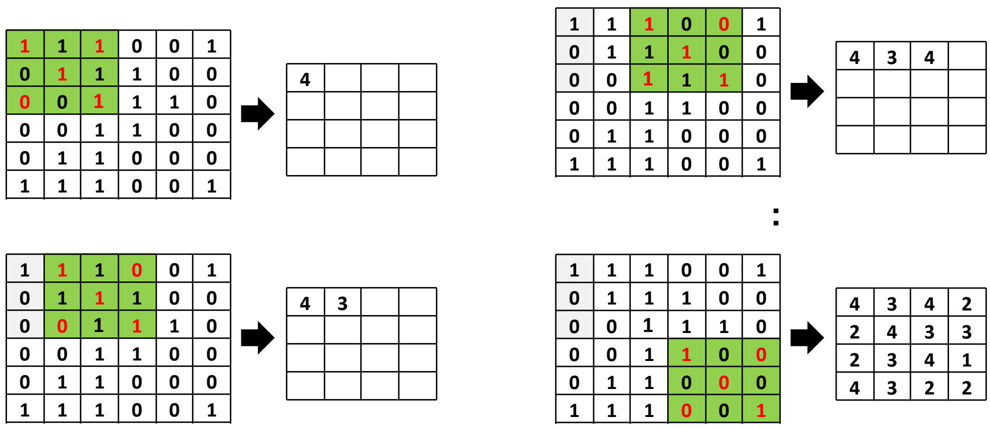


Figure 3 Convolution of a 6x6 Colored Image

2.3 Max Pooling Layers

CNN provide the local and global layers that combine the convolutions within the pool. The combination of region within convolution can reduce the spatial size of the representation that makes CNN totally less computation than MLP. There are various non-linear functions to the pooling layers. Most CNN-based computer models use “Max Pooling Layers: MPL” instead of “Average Pooling Layers: APL”. MPL is better in performance than APL because the finding of the max value is totally faster than the average. For example, a max pooling of a 4x4 convolution can be shown in figure 4.

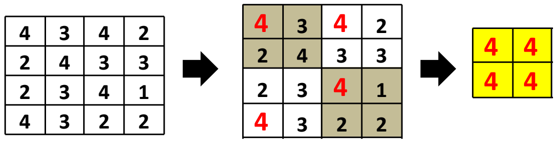


Figure 4 Max Pooling Layers of a 4x4 Convolution

2.4 Rectified Linear Units with Output Classes

In order to fully connected layers of the CNN-based computer model, the "Rectified Linear Units (RELU)" can be seen as the decision function (also called "activation function") to select the "neurons ($a(x_i)$)" of computer model. The RELU considers the neuron from its fully connected layers to another one. The RELU using absolute of hyperbolic tangent function is well-known to train with several times faster than other functions (that can be defined as $RELU(x_i) = |\tanh(x_i)|$).

To build neurons ($a(x_i)$), the CNN-based computer model can be trained from set of 480 recipe-tagged images (X) (where $X = \{x_1, x_2, x_3, ..., x_{480}\}$) and their output classes (Y) (where $Y = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\}$). The output classes are according to 7 Thai-desserts as Khanom Krok, Coco Ice-cream, Fruity Luk Chub, Woon Bai Toey, Tub Tim Krob, Luem Gluen and Bua Loy. The training set of CNN can be done using cross-entropy loss, by (1).

$$Model(X,Y) = -\left(\frac{\sum_{i=1}^{480} (\ln a(x_i) + (1 - y_i)\ln(1 - a(x_i)))}{480} \right) \quad (1)$$

3. Name and Recipe Estimation of Dessert Images

Our CNN-based computer model is built from 480 recipe-tagged images that cover 7 Thai-desserts as as Khanom Krok, Coco Ice-cream, Fruity Luk Chub, Woon Bai Toey, Tub

Tim Krob, Luem Gluen and Bua Loy. The unknown dessert image is estimated (or tagged) to search for its name and recipes by the model as shown in figure 5.

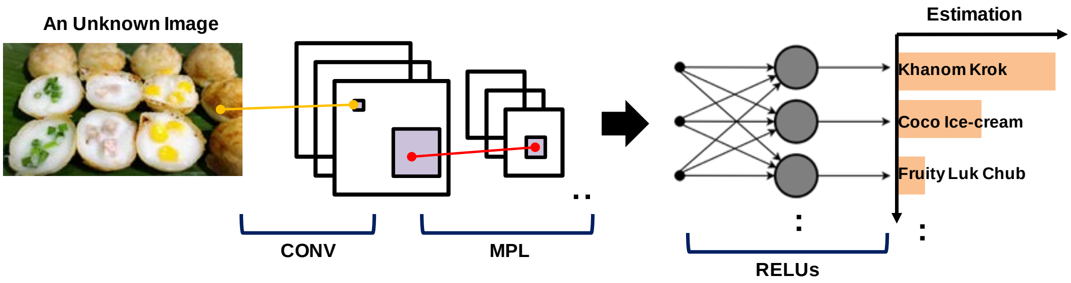


Figure 5 Name and Recipe Estimation of an Unknown Image

To estimate the name and recipes, the probability of output classes has been computed by (2).

$$P(y = j \mid x, w_1, w_2 ... w_7, b_1, b_2 ... b_7) = \frac{e^{w_j x + b_j}}{\sum_{k=1}^{7} e^{w_k x + b_k}} \quad (2)$$

where x is the result from RELUs, w_j and b_j are the weight and bias values in the $j-th$ neuron, and y is any output class (from 7 classes). The estimation is finally done by the highest probability of output class ($\hat{y}$) as (3).

$$\hat{y} = \arg_j \max\left(P(y = j \mid x, w_1, w_2 ... w_7, b_1, b_2 ... b_7)\right) \quad (3)$$

With this in mind, our proposed "Name and Recipe Estimation of Thai-desserts beyond Image Tagging" is evaluated by the accuracy criteria that can be computed by (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where TP (True Positive) is the dessert "A" and our model estimates correctly as "A", TN (True Negative) is not the dessert "A" and our model estimates correctly as not "A", FP

(False Positive) is not the dessert "A" and our model estimates wrongly as "A" and FN (False Negative) is the dessert "A" and our model estimates wrongly as not "A". The results are shown in Table 1.

Table 1 The accuracy of Thai-dessert Recognition

Thai-dessert	Accuracy (%)
Khanom Krok (Thai: ขนมครก)	82
Coco Ice-cream (Thai: ไอศกรีมกะทิสด)	95
Fruity Luk Chub (Thai: ลูกชุบผลไม้)	93
Woon Bai Toey (Thai: วุ้นใบเตย)	87
Tub Tim Krob (Thai: ทับทิมกรอบ)	78
Luem Gluen (Thai: ลืมกลืน)	85
Bua Loy (Thai: บัวลอย)	81
Average	**86**

Since Convolutional Neural Network (CNN) is a successfully well-known image tagging that totally provides high accuracy. As our CNN-based computer model for estimating the name and recipe of unknown Thai-dessert images are generally done over 85%. Having said that, the similarity between "Tub Tim Krob" and "Bua Loy" is sometimes difficult to distinguish and the accuracy becomes lower than other Thai-desserts. The "Coco Ice-cream" and "Fruity Luk Chub" provides the highest recognition accuracy because of their uniqueness(s).

4. Conclusions

Since Thai-desserts mean the "Thai folk wisdom (Thai: ภูมิปัญญาไทย)". In 2015, The Food Republic recommended the 7 Thai-deserts. However, the inconvenience of finding these Thai names and recipes means the international outsiders do not dare to have these Thai-desserts. Hence, this paper proposes a novel combination between "Artificial Intelligence" and "Computer Vision" to build a CNN-based computer model for tagging the name and recipe of unknown Thai-dessert image. The computer model is built from 480 recipe-tagged images that cover those 7 Thai-desserts. From the experimental results, our computer model provides the accuracy higher than 85%. For future work, this research can be used the newest “Image-to-image translation techniques” [31] of computer vision that will provide better speed and higher correctness.

References

[1] CNN.com. Khanom time! Thailand’s inimitable snacks and desserts [Internet]. 2010 [cited 2017 Dec 5]. Available from: http://travel.cnn.com/bangkok/eat/guide-thailands-colorful-snacks-and-desserts-105745/.

[2] Foodrepublic.com. 7 Thai Desserts You Didn’t Know Existed [Internet]. 2015 [cited 2017 Dec 5]. Available from: http://www.foodrepublic.com/2015/01/15/7-thai-desserts-you-didnt-know-existed/.

[3] Mthai.com. Thai Wisdom: Wrappers in Thai Dessert [Internet]. 2013 [cited 2017 Dec 5]. Available from: https://en.mthai.com/thaiculture/1321.html.

[4] Theguardian.com. Bangkok’s best street food: a guide to dishes and districts [Internet]. 2015 [cited 2017 Dec 5]. Available from: https://www.theguardian.com/travel/2015/nov/25/bangkok-best-street-food-guide-dishes-districts-thailand.

[5] Olaode AA, Naghdy G and Todd CA. Bag-of-Visual Words Codebook Development for the Semantic Content Based Annotation of Images. In: The 11th IEEE International Conference on Signal-Image Technology & Internet-Based Systems; 2015 Nov 23; Bangkok, Thailand. p.7-14.

[6] Olaode AA, Naghdy G and Todd CA. Efficient region of interest detection using blind image division. In: The 2015 IEEE Signal Processing Symposium; 2015 Jun 10; Debe, Poland. p.1-6.

[7] Soimart L and Pongcharoen P. Multi-row Machine Layout Design using Artificial Bee Colony. In: The International Conference on Economics and Business Information; 2011 May 1; Bangkok, Thailand.

[8] Soimart L and Mookdarsanit P. An Admission Recommendation of High-school Students using Apriori Algorithm. In: The 6th International Conference on Sciences and Social Science; 2016; Mahasarakham, Thailand.

[9] Mookdarsanit P and Ketcham M. Image Location Estimation of well-known Places from Multi-source based Information. In: The 11th International Symposium on Natural Language Processing; Ayutthaya, Thailand.

[10] Mookdarsanit P and Rattanasiriwongwut M. Location Estimation of a Photo: A Geo-signature MapReduce Workflow. Engineering Journal 2017; 21: 295-308.

[11] Huang Q, Jia CK, Zhang X and Ye Y. Learning Discriminative Subspace Models for Weakly Supervised Face Detection. IEEE Transactions on Industrial Informatics 2017;13:2956-64.

[12] Soimart L and Mookdarsanit P. Gender Estimation of a Portrait: Asian Facial-significance Framework. In: The 6th International Conference on Sciences and Social Science; 2016; Mahasarakham, Thailand.

[13] Mookdarsanit P and Rattanasiriwongwut M. GPS Determination of Thai-temple Arts from a Single Photo. In: The 11th International Conference on Applied Computer Technology and Information Systems; 2017; Bangkok, Thailand. p.42-7.

[14] Mookdarsanit P and Rattanasiriwongwut M. MONTEAN Framework: A Magnificent Outstanding Native-Thai and Ecclesiastical Art Network. International Journal of Applied Computer Technology and Information Systems 2017;6:17-22.

[15] Lv J, Shao X, Xing J, Cheng C, Zhou X. A Deep Regression Architecture with Two-Stage Re-initialization for High Performance Facial Landmark Detection. In: The 2017 IEEE Conference on Computer Vision and Pattern Recognition; 2017; Honolulu. p.3691-700.

[16] Soimart L and Ketcham M. An Efficient Algorithm for Earth Surface Interpretation from Satellite Imagery. Engineering Journal 2016;20:215-28.

[17] Soimart L and Ketcham M. Hybrid of Pixel-based and Region-based Segmentation for Geology Exploration from Multi-spectral Remote Sensing. In: The 11th International Symposium on Natural Language Processing; 2016; Ayutthaya, Thailand.

[18] Elmannai H, Loghmari MA and Naceur MS. Nonlinear separation source and parameterized feature fusion for satelite image patch exemplar. In: The 2015 IEEE International Geoscience and Remote Sensing Symposium; 2015; Milan, Italy. p.405-8.

[19] Soimart L and Mookdarsanit P. Name with GPS Auto-tagging of Thai-tourist Attractions from An Image. In: The 2017 Technology Innovation Management and Engineering Science International Conference; 2017; Nakhon Pathom, Thailand.

[20] Zhao Z, Ma H and You S. Single Image Action Recognition Using Semantic Body Part Actions. In: The 2017 IEEE International Conference on Computer Vision; 2017; Venice, Italy. p.3411-9.

[21] Mookdarsanit P and Mookdarsanit L. An Automatic Image Tagging of Thai Dance's Gestures. In: Joint Conference on ACTIS & NCOBA; 2018 Jan; Ayutthaya, Thailand. p.76-80.

[22] Mookdarsanit P and Mookdarsanit L. A Content-based Image Retrieval of Muay-Thai Folklores by Salient Region Matching. International Journal of Applied Computer Technology and Information Systems 2018;7:21-6.

[23] Ciocca G, Napoletano P and Schettini R. Food Recognition: A New Dataset, Experiments, and Results. IEEE Journal of Biomedical and Health Informatics 2017;21: 588-98.

[24] Herranz L, Jiang S and Xu R. Modeling Restaurant Context for Food Recognition. IEEE Transactions on Multimedia 2017;19:430-40.

[25] Soimart L and Mookdarsanit P. Ingredients estimation and recommendation of Thai-foods. SNRU Journal of Science and Technology 2017;9:509-20.

[26] Termritthikun C, Muneesawang P and Kanprachar S. NU-INNET: Thai Food Image Recognition using Convolutional Neural Networks on Smartphone. Journal of Telecommunication Electronic and Computer Engineering 2017;9:2-6.

[27] Sun T, Wang Y, Yang J and Hu X. Convolution Neural Networks With Two Pathways for Image Style Recognition. IEEE Transactions on Image Processing 2017;26:4102-13.

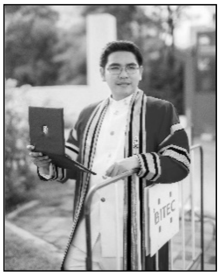
[28] Hsu CC and Lin CW. CNN-Based Joint Clustering and Representation Learning with Feature Drift Compensation for Large-Scale Image Data. IEEE Transactions on Multimedia 2018;20:421-9.

[29] Soimart L and Ketcham M. The Segmentation of Satellite Image Using Transport Mean-shift Algorithm. In: The 13th International Conference on IT Applications and Management; 2015; Phitsanulok, Thailand. p.124-8.

[30] Mookdarsanit P, Soimart L, Hnoohom N and Ketcham M. Detecting Image Forgery Using XOR and Determinant of Pixels for Image Forensics. In: The 11th IEEE International Conference on Signal-Image Technology & Internet-Based Systems; 2015 Nov 23; Bangkok, Thailand. p.613-6.

[31] Zhu JY, Park T, Isola P and Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: The 2017 IEEE International Conference on Computer Vision; 2017; Venice, Italy. p.2242-51.

Author's Profile



Pakpoom Mookdarsanit (Ph.D.) is currently a full-time lecturer at Computer Science Department and the chief of Computer Field Experience Center at Chandrakasem Rajabhat University. He received his B.Eng. in Computer Engineering from Mahidol University and M.S. in Computer Science from Kasetsart University in 2010 and 2013, respectively. In 2017, he received his Ph.D. in Information Technology from King Mongkut's University of Technology North Bangkok. His research involves image processing and machine learning. E-Mail: pakpoom.m@chandra.ac.th

Lawankorn (Soimart) Mookdarsanit (Ph.D.) received her B.BA. in Business Computer (with high honors) and M.S. in Information Technology from Naresuan University in 2009 and 2011, respectively. And the Ph.D. in Information Technology was done in 2016 from King Mongkut's University of Technology North Bangkok. Currently, she is a lecturer and the chief of Business Computer Department, Chandrakasem Rajabhat University. Her researches are computer vision, pattern recognition and data mining. E-Mail: lawankorn.s@chandra.ac.th