Azerbaijan Journal of High Performance Computing, Vol 3, Issue 1, 2020, pp. 75-93
https://doi.org/10.32010/26166127.2020.3.1.75.93

ThaiWrittenNet: Thai Handwritten Script Recognition using Deep Neural Networks

Pakpoom Mookdarsanit, Lawankorn Mookdarsanit

Chandrakasem Rajabhat University, Bangkok, Thailand, pakpoom.m@chandra.ac.th, lawankorn.s@chandra.ac.th

*Correspondence: Pakpoom Mookdarsanit, Chandrakasem Rajabhat University, Bangkok, Thailand, pakpoom.m@chandra.ac.th

Abstract

Thai is a non-tonal language usage for 70 million speakers in Thailand. A variety of Thai handwrit-ten styles has been a challenge in handwriting recognition. In this paper, we propose a novel "ThaiWrittenNet" based on Convolutional Neural Network (ConvNet or CNN) with a cutout to identify the handwritten recognitions. Deep Belief Network (DBN) is also combined with Con-vNet to reduce network complexity. From the results, ThaiWrittenNet outperforms the flat Con-vNet and other handcrafted features with traditional machine learning algorithms. It appears that DBN helps ConvNet to improve the accuracy of Thai-handwritten recognition.

Keyword: Handwriting recognition, Convolutional neural network, Deep belief network, Thai handwriting recognition

1. Introduction

Intelligent Thai handwritten script recognition refers to the optical scanning of a handwritten image as an input and interpreting it into textual information as an output (Surinta, Karaaba, Schomaker & Wiering, 2015). The first handwritten interpretation was addressed by a group of researchers from the Center of Excellence for Document Analysis and Recognition (CEDAR); and implemented by English (simple A-Z) characters as a form of sketch recognition (Srihari & Kuebert, 1997). In contrast, the printed character recognition can be seen as a problem of image segmentation (Liu, Bober & Kittlet, 2019) to divide all pixels (Soimart & Ketcham, 2016b) Into the object or background sets (Soimart & Ketcham, 2015). As the world has more than 7,000 spoken languages (Ager, 2020), e.g., Thai, Laos, Khmer, Burmese, Javanese, Bangla, Chinese, Japanese, Hindi, Arabic, etc., the handwritten script recognition consequently have an enormous variety of hand-writing styles, sizes, and shapes that is still a hard and open problem (Alom, Sidike, Taha & Asari, 2017), unlike printed character recognition (Ismayilov & Mammadov, 2019; Emsawas & Kijsirikul, 2016; Chaiwatanaphan, Pluempitiwiriyawej & Wangsiripitak, 2017). Furthermore, various writers' language makes different writing identities: separated-characters or connected-characters (Pornpanomchai, Wongsawangtham, Jeungudomporn & Chatsumpun, 2011), recognized by one challenge of Natural Language Processing (NLP) hot areas.

1.1. Thai and computational linguistics

Longer than 720 years, Thai has been a spoken and written language (Satienkoses, 1981). in Thailand (or Siam) since Sukhothai (Thai: กรุงสุโขทัย), Ayutthaya (Thai: กรุงศรีอยุธยา) until Rattanakosin (Thai: กรุงรัตนโกสินทร์) era. From the historical heritage, the old Thai scripts were inscribed on the memorial stones (Thai: ศิลาจารึก) by King Ramkhamhaeng (Thai: พ่อขุนรามคำแหง of Sukhothai (Inthajakra, Prachyapruit & Chantavanich, 2016) that was officially announced as one of memory of the world by UNESCO in 2003. Thai is one of Kradai (Thai: ขร้า-ไท) language family that most Thai words and/or vocabularies inherit from Sanskrit, Pali, Khmer, and Mon (Satienkoses, 1981). Thai was also used to collect the literature doctrines in Buddhist Scriptures (Thai: พระไตรปิฎก). Up until now, almost 70 million speakers use Thai (either writing or typing) as an official language (World Bank, 2018) in their daily life, such as an envelope, health check-up form, official document, individual tax, Buddhism quotes, and many more.

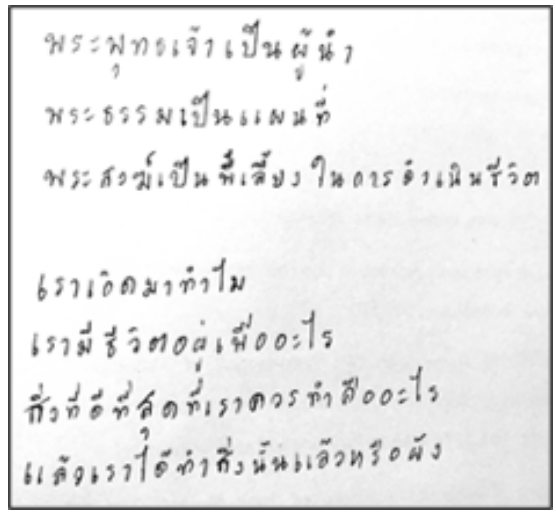

Fig.1. A personal Thai handwritten note on the remembrance of Phra Phrom Mangkhalachan (Thai: พระพรหมมังคลาจารย์)'s Buddhism preachment (Wat Chonprathan Rangsarit, 2001).

Linguistically, Thai is one of the tonal languages that one pronunciation in different tones has various meanings as one of the challenges in 5G testbeds for Thai voice and tone (Daengsi & Wuttidittachotti, 2019). Like a word "Pa" in Thai has five different tones: the first tone (Thai: ปา, v.) means throwing something away, the second tone (Thai: ป่า, n.) as forest, the third tone (Thai: ป้า, n.) as aunt, the fourth (Thai: ป๊า, n.) and fifth tone (Thai: ป๋า, n.) as father, respectively. Unlike English written style, a Thai sentence or phrase has no space between 2 words (Haruechaiyasak, Kongyoung & Dailey, 2008) that needs some complex algorithms for word

segmentation (Klahan, Pannoi, Uewichitrapochana & Wiangsripanawan, 2018). Moreover, it is one of the challenges in Thai Natural Language Processing (Thai-NLP) (Koanantakool, Karoonboonyanan & Wutiwiwatchai, 2009). The categorization of Thai-NLP (Sornlertlamvanich, Potipiti, Wutiwiwatchai & Mittrapiyanuruk, 2000) researches are Thai NLP understanding (Nomponkrang & Sanrach, 2016). Thai word segmentation (Theeramunkong & Tanhermhong)., statistical machine translation (Lyons, 2016), Thai sentiment analysis (Haruechaiyasak, Kongthon, Palingoon & Trakultaweekoon, 2013) and Thai handwritten recognition (Surinta & Nitwuwat, 2006). According to the big data era, most of all, statistically-based embedding methods have been changed into deep learning, which has many advanced attention techniques (Raghu & Schmidt, 2020) for Thai-NLP: transfer adaptation learning, deep reinforcement learning, augmentation, semi-supervision, etc. Some Thai-NLP papers based on deep learning are available: Thai bully detection (Mookdarsanit & Mookdarsanit, 2019), part of speech tagging (Boonkwan & Supnithi, 2017) and Thai word segmentation (Lapjaturapit, Viriyayudhakom & Theeramunkong, 2018). A popular sequence-to-sequence self-attention such as TRANSFORMER or BERT can be applied for the languages with tonal markers like Thai. Furthermore, there are many other hot areas (Torfi, Shivani, Keneshloo, Tavvaf & Fox, 2020) for Thai-NLP: image captioning, sentiment analysis, visual/textual question answering, document summarization, and dialogue system. Human resource (HR) intelligence with Thai-NLP (Mookdarsanit & Mookdarsanit, 2020b). Is still a developing technology for Thai organizations. Likewise, Thai writing with 44 Thai characters, 32 vowels, five tones, and 10 Thai numerals, coupled with hugely-different written styles, are still opened for deep learning (Zou, Shi, Guo & Ye, 2019), like Convolutional Neural Network (ConvNet or CNN).

1.2. The proposed ThaiWrittenNet

Although Thai handwritten script recognition can be seen as a form of character image recognition, the handwriting has more challenges in a large variety of Thai handwritten styles by different writers that affect the recognition accuracy. Previous handwriting recognition researches in other languages are available and can be classified into 2 ages (Zheng, Yang & Tian, 2017): traditional machine learning (e.g., k-NN, MLP and SVM) and deep learning (e.g., ConvNet or CNN). In 2012, AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) showed the ConvNet (Liu, Ouyang, Wang, Fieguth, Chen, Liu & Pietikäinen, 2019) over handcrafted features with traditional machine learning (Olaode, Naghdy & Todd, 2014) that changed the world of object recognition (Alom, Taha, Yakopcic, Westberg, Sidike, Nasrin, Esesn, Awwal & Asari, 2018). In this paper, we propose a novel ThaiWrittenNet deep learning that combines the Convolutional Neural Network (ConvNet) and Deep Belief Network (DBN). From the experiment, DBN can help ConvNet to reduce the complexity but higher accuracy for Thai-handwritten recognition. Moreover, ConvNet outperforms traditional machine

learning.

This paper is organized into 6 parts. Handcrafted feature extraction and traditional supervised models are described in parts 2 and 3. Part 4 deeply talks about the convolutional neural network. Experimental results and discussion is in part 5. Finally, part 6 is the conclusion.

2. Handcrafted feature extraction

Although deep learning has beat traditional handcrafted features (Zheng, Yang & Tian, 2017), some papers are based on handcrafted feature extraction (Olaode & Naghdy, 2020) with some evolutionary optimization (Soimart & Pongcharoen, 2011) techniques. A Thai-handwritten image is technically extracted into features in terms of a vector (a.k.a. word or codebook) prior to being classified by traditional machine learning. The handcrafted feature extraction (Mookdarsanit & Ketcham, 2016) for Thai-handwriting consists of Scale Invariant Feature Transform (SIFT), Speed Up Robust Feature (SURF), and Histogram of Gradients (HoG).

2.1. Scale Invariant Feature Transform (SIFT)

Lowe introduced scale Invariant Feature Transform (SIFT) in 2004 (Lowe, 2004). The SIFT has either a detector or a descriptor. In short, the detector is proposed to find an exciting feature (a.k.a. key-points) within a handwritten image. Later, those key-points are quantized into words and kept in a vector (a.k.a. word or codebook) by the descriptor. Many handwritten images are detected and described by SIFT as many vectors stored in a bag of words (Mookdarsanit & Rattanasiriwongwut, 2017a). The SIFT in terms of Laplacian ($L(x,y,\sigma)$) as a blob detector can be computed by the convolution (Mookdarsanit & Mookdarsanit, 2018b). Between a Thai-handwritten image ($I_{ThaiWritten}(x,y)$) and Gaussian scale kernel ($G(x,y,\sigma)$) by

$$L(x,y,\sigma) = I_{ThaiWritten}(x,y) * Gauss(x,y,\sigma), \quad (1)$$

where x, y is the position of pixel intensity of a Thai-handwritten image, σ is the width of Gaussian kernel and $Gauss(x,y,\sigma) = \frac{1}{2\pi^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}$.

Concretely, the full Laplacian can be approximately computed by the Difference of Gaussian (Zheng, Yang & Tian, 2017), in

$$D(x,y,\sigma_{\nabla ij}) = L(x,y,\sigma_i) - L(x,y,\sigma_j). \quad (2)$$

Then, the horizontal and vertical SIFT gradient (G_{SIFT}) is computed by Laplacian in x and y neighborhood pixels (Soimart & Ketcham, 2016a) by

$$G_{SIFT} = \begin{cases} L(x+1,y,\sigma) - L(x-1,y,\sigma), & if \quad x-axis, \\ L(x,y+1,\sigma) - L(x,y-1,\sigma), & if \quad y-axis. \end{cases} \quad (3)$$

And the magnitude $M_{SIFT}(\bullet)$ and orientation $\theta_{SIFT}(\bullet)$ of the SIFT gradient can be computed by

$$M_{SIFT}(x,y) = \sqrt{(G_{SIFT\ in\ x})^2 + (G_{SIFT\ in\ y})^2} \tag{4}$$

and

$$\theta_{SIFT}(x,y) = \tan^{-1}\left(\frac{G_{SIFT\ in\ y}}{G_{SIFT\ in\ x}}\right). \tag{5}$$

Since the handwritten image is divided into 16 blocks. Each block is plotted on a histogram. The orientation histogram takes 8 bins for all possible 360 directions (each of them as 45), which results in 128 dimensions of a vector (a.k.a. 128-D SIFT) for different scales (Zheng, Yang & Tian, 2017). For higher speed, there are so many SIFT dimensions (Olaode, Naghdy & Todd, 2014) such as PCA-SIFT, 64-D SIFT, and SURF.

2.2. Speed Up Robust Feature (SURF)

In 2006, Speed Up Robust Feature (SURF) was designed to solve the SIFT complexity – as a modified version of SIFT (Bay, Tuytelaars & Gool, 2006). SURF only has 64 dimensions (Mookdarsanit & Mookdarsanit, 2018c), which is both detection and description, less complexity than SIFT. Firstly, the basic image is computed through the all handwritten image's pixels by

$$Integral_{IMG}(x_i, y_j) = \sum_{u=0}^{i}\sum_{v=0}^{j} P_{(x_u, y_v)}, \tag{6}$$

where x_i, y_j refer to the position of the pixel of a handwritten image, $P_{(x_u,y_v)}$ is any positions before $P_{(x_i,y_j)}$, like

$$P_{(0,0)}, P_{(0,1)}, P_{(0,2)}, P_{(0,3)}, .., P_{(1,0)}, P_{(1,1)}, P_{(1,2)}, P_{(1,3)}; ..., P_{(2,0)}, P_{(2,1)}, P_{(2,2)}, P_{(2,3)}, ..., P_{(x_i,y_j)}.$$

Second, the Gaussian Second-Order Derivatives (as well as SIFT) in 3 dimensions: x^2, y^2, and xy (Mookdarsanit & Mookdarsanit, 2018c) are computed by

$$\begin{aligned}
\frac{\partial^2}{\partial x^2} Gauss(x,y,\sigma) &= \frac{\partial^2}{\partial x^2}\left(\frac{1}{2\pi\sigma^2} \times e^{\frac{x^2+y^2}{2\sigma^2}}\right), \\
\frac{\partial^2}{\partial y^2} Gauss(x,y,\sigma) &= \frac{\partial^2}{\partial y^2}\left(\frac{1}{2\pi\sigma^2} \times e^{\frac{x^2+y^2}{2\sigma^2}}\right), \\
\frac{\partial^2}{\partial x \partial y} Gauss(x,y,\sigma) &= \frac{\partial^2}{\partial x \partial y}\left(\frac{1}{2\pi\sigma^2} \times e^{\frac{x^2+y^2}{2\sigma^2}}\right).
\end{aligned} \tag{7}$$

Third, a difference between (5) and (6) is computed, called the difference between Gaussian Second-Order Derivatives and Integral Image (Mookdarsanit & Mookdarsanit, 2018c) in 3 dimensions: $D_{xx}(x,y,\sigma)$ $D_{yy}(x,y,\sigma)$ and $D_{xy}(x,y,\sigma)$

$$\begin{aligned} D_{xx}(x,y,\sigma) &= \left| \frac{\partial^2}{\partial x^2} Gauss(x,y,\sigma) - Integral_{IMG}(x_i, y_j) \right|, \\ D_{yy}(x,y,\sigma) &= \left| \frac{\partial^2}{\partial y^2} Gauss(x,y,\sigma) - Integral_{IMG}(x_i, y_j) \right|, \\ D_{xy}(x,y,\sigma) &= \left| \frac{\partial^2}{\partial x \partial y} Gauss(x,y,\sigma) - Integral_{IMG}(x_i, y_j) \right|. \end{aligned} \tag{8}$$

Image $\{D_{xx}(x,y,\sigma),\ D_{yy}(x,y,\sigma),\ D_{xy}(x,y,\sigma)\}$ are stored in terms of a Hessian Matrix (Bay, Tuytelaars & Gool, 2006)

$$H(x,y,\sigma) = \begin{bmatrix} D_{xx}(x,y,\sigma) & D_{xy}(x,y,\sigma) \\ D_{xy}(x,y,\sigma) & D_{yy}(x,y,\sigma) \end{bmatrix}. \tag{9}$$

Fifth, the determinant (Mookdarsanit, Soimart, Ketcham & Hnoohom, 2015) of $H(x,y,\sigma)$ is computed by

$$\det(H(x,y,\sigma)) = \left(D_{xx}(x,y,\sigma) \times D_{yy}(x,y,\sigma)\right) - D_{xy}{}^2(x,y,\sigma). \tag{10}$$

After that, the magnitude and orientation of SURF is separated into x and y-axis (Bay, Tuytelaars & Gool, 2006), where $x = \{-\cos\theta, \cos\theta\}, y = \{-\sin\theta, \sin\theta\}$ $0 \le \theta \le 360$ and, computed by

$$\begin{aligned} M_{SURF}(x) &= \sum_{i=0}^{\det(H(x,y,\sigma))_u} \left(\det(H(x,y,\sigma))_i \times \left|\cos\theta_j\right|\right), \\ M_{SURF}(y) &= \sum_{i=0}^{\det(H(x,y,\sigma))_u} \left(\det(H(x,y,\sigma))_i \times \left|\sin\theta_j\right|\right) \end{aligned} \tag{11}$$

and

$$\begin{aligned} \theta_{SURF}(x) &= \sum_{j=1}^{n(\theta_j)} n(\det(H(x,y,\sigma))_i \mid \pm\cos\theta_j), \\ \theta_{SURF}(y) &= \sum_{j=1}^{n(\theta_j)} n(\det(H(x,y,\sigma))_i \mid \pm\sin\theta_j). \end{aligned} \tag{12}$$

2.3. Histogram of Gradient (HoG)

Histogram of the gradient (HoG) or Dense-SIFT (Olaode, Naghdy & Todd, 2014) is an image descriptor by counting the occurrence of gradient density and orientation (Zheng, Yang & Tian, 2017). HoG was introduced in 2005 (Dalal & Triggs, 2005). Later, a well-known Support Vector Machine (SVM) learning model that works efficiently with HoG for pedestrian and face detection (Soimart & Mookdarsanit, 2016a). For the preprocessing, the handwritten image size is set into ratio 1:2, e.g., 100×200, 128×256, or 1000×2000. First, the handwritten image is convoluted by x and y Sobel filtering windows, also called the HoG gradient ($G_{HoG}(\bullet)$), by

$$\begin{aligned} G_{HoG}(x) &= I_{ThaiWritten}(x,y) \bullet [-1\ 0\ 1], \\ G_{HoG}(y) &= I_{ThaiWritten}(x,y) \bullet [-1\ 0\ 1]^T, \end{aligned} \tag{13}$$

where x, y refer to the position of pixel intensity of a Thai-handwritten image, $[-1\ 0\ 1]$ is x-axis Sobel window and $[-1\ 0\ 1]^T$ is a transpose of $[-1\ 0\ 1]$ in y-axis.

Then, the magnitude $M_{HoG}(\bullet)$ and orientation $\theta_{HoG}(\bullet)$ of the HoG gradient (Mookdarsanit & Rattanasiriwongwut, 2017c). can be computed by (as well as those of SIFT)

$$M_{HoG}(x,y) = \sqrt{{G_{HoG}}^2(x) + {G_{HoG}}^2(y)} \tag{14}$$

and

$$\theta_{HoG}(x,y) = \tan^{-1}\left(\frac{G_{HoG}(y)}{G_{HoG}(x)}\right). \tag{15}$$

Next, the image is divided into cells that each cell has size as 8x8 pixels. In each cell, the $M_{HoG}(x,y)$ is plotted in the 9-bin $\theta_{HoG}(x,y)$ graph, called histogram of gradient (HoG). After that, the neighbor 4 cells are grouped into the block as 16x16 block normalization (Dalal & Triggs, 2005). Finally, each block has 4 HoGs, which means each block has 9x4=36 dimensions in the vector (Olaode, Naghdy & Todd, 2014).

3. Traditional supervised models

From the output of feature extraction (SIFT, SURF, or HoG), all features (a.k.a. interesting points) from an image (Olaode & Naghdy, 2020) are stored within a numerical vector representation (Mookdarsanit & Rattanasiriwongwut, 2017b). Each Thai-handwritten recognition image is represented by a row of the vector (a.k.a. Bag of words) that is used to train and/or test in the supervised learning models, as shown in Figure 2.

3.1. K-nearest neighbor (k-NN)

K-nearest neighbor (k-NN) is instance-based supervised learning, based on statistical estimation. Since each Thai-handwritten recognition image is stored in a row of the vector (Rathi, Pandey, Chaturvedi & Jangid, 2012), all rows are kept as a big

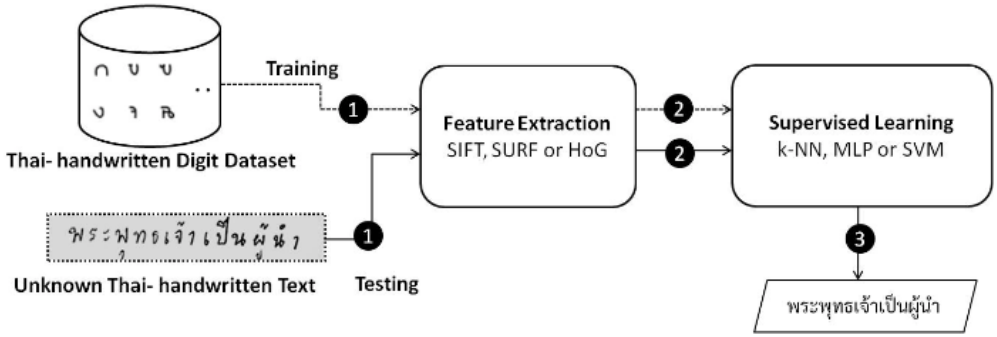

Fig. 2. Supervised learning: training and testing

functions (a.k.a. lazy learning). For testing, the minimum distance $\min(D(\bullet))$ between an unknown recognition ($V_{unknown}$) and some Thai-handwritten recognition images ($V_{TH\ digit}$ (s)) are compared by Euclidean distance

$$D(V_{unknown}, V_{TH\ digit}) = \sqrt{\sum_{i=1}^{n}\left(v_{unknown\ i} - v_{TH\ digit\ i}\right)^2}, \quad (16)$$

where $v_{unknown\ i} \in V_{unknown}$, $v_{TH\ digit\ i} \in V_{TH\ digit}$ and n is the number of rows within a vector.

3.2. Multi-layer perceptron (MLP)

Multi-layer perceptron (MLP) is a classical deep learning function that maps all inputs into output without reasoning (Rumelhart & McClelland, 1987). As for Figure 3, the basic unit perceptron receives imaging data (such a Thai-handwritten recognition image) as input nodes $[x_1, x_2, x_3, x_4, ..., x_{length\ of\ row}]$ from a row of the vector (Soimart & Mookdarsanit, 2017a). The internal parameters consist of synaptic weights $[w_{k1}, w_{k2}, w_{k3}, ..., w_{km}]$ and biases (bk) learned during training (a.k.a. the linear combination of input signals with a bias).

The output (yk) is produced by activation function ($f(\bullet)$) with the effect of an affine transformation. The unit perceptron can be mathematically represented by

$$y_k = f\left(\sum_{j=1}^{m}\left(w_{kj}x_j\right) + b_k\right) = f\left(W^k x_j + b_k\right), \quad (17)$$

where W^k refers to $\sum_{j=1}^{m}\left(w_{kj}x_j\right)$.

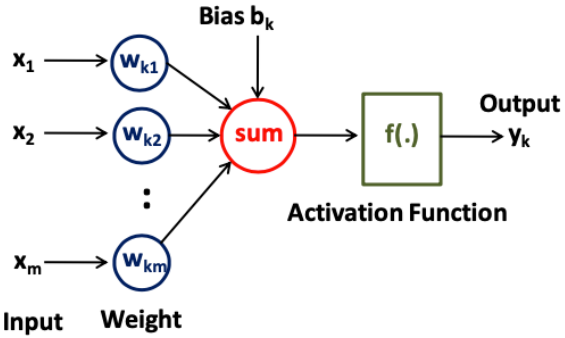

Fig.3. A unit perceptron

MLP has many hidden layers with multiple hidden nodes with more parameters (Soimart & Mookdarsanit, 2017a) for training, as shown in Figure 4.

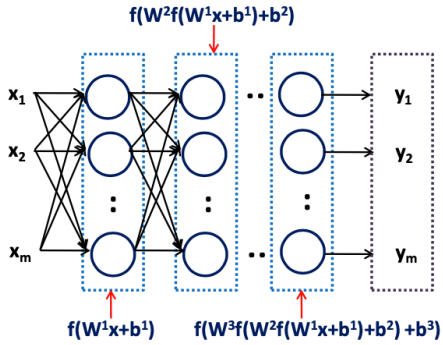

Fig. 4. Multi-layer perceptron

3.3. Support vector machine (SVM)

Support vector machine (SVM) is applied to many object recognition and image classification applications (Soimart & Mookdarsanit, 2017b). SVM's concept is to find the optimal hyperplane (Cortes & Vapnik, 1995) that has the maximum distance of 2 closet data points (a.k.a. support vectors) between 2 classes. SVM is based on a linear binary classifier. The decision function is given by

$$f(x) = sign(w^T x + b), \tag{18}$$

where w is a transpose of the weight vector, b is the bias that is computed by cost function

$$J(w,\xi) = \frac{1}{2} w^T w + C\sum_{i=1}^{n} \xi_i, \tag{19}$$

where C is a trade-off between training error and generalization, ξ_i is the i-th slack variable to tolerate the errors and be minimized. Given hyperplane $w^T x + b = 0$ with the splitting hyperplane obtains the max distance between closet positives $w^T x + b = +1$ and negatives $w^T x + b = -1$. Moreover, the non-linear kernel function is radial basis function (RBF), to compute the similarity between 2 vectors, where γ is RBF kernel parameters

$$K_{RBF}(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right). \quad (20)$$

4. Convolutional neural network (ConvNet)

Convolutional neural network (ConvNet or CNN) is deep learning proposed for image data that was firstly introduced in 1980 (Fukushima, 1980). In the 1990s, the ConvNet with gradient-based learning showed successful recognition, traffic signs, and handwritten recognition (LeCun, Bottou, Bengio & Haffner, 1998). AlexNet (Krizhevsky, Sutskever, & Hinton, 2012) obtained great performance recognition by ConvNet with ImageNet in 2012. Afterward, ConvNet was proven to be more accurate than a handcraft feature with traditional machine learning. On the dark side, ConvNet was used to train the spamming bots to recognize those reCaptcha images and break the human verification (Mookdarsanit & Mookdarsanit, 2020a) against authentication (Soimart & Mookdarsanit, 2016b) mechanism that finally made the sever-side system processed a large number of junk jobs as the concurrent workloads (Mookdarsanit & Gertphol, 2013). ConvNet was designed to have highly optimized structures (Mookdarsanit & Mookdarsanit, 2019a) to learn the extraction and abstraction of 2D features. Especially, the shape variations are solved by the Max-pooling layer (Mookdarsanit & Mookdarsanit, 2020).

Most ConvNet is trained by gradient-based learning that suffers less from the diminishing gradient problem. ConvNet has 2 essential parts (Mookdarsanit & Mookdarsanit, 2018d): feature extraction (consists of convolution in even-numbered and max-pooling in odd-numbered layers) and classification. The output of convolution and max-pooling is called feature mapping (Mookdarsanit, 2020). Each node of the convolution layer extracts features from the input Thai-handwritten image by a convolution operation. Max-pooling layer abstracts those features by average or propagating operation over the input nodes.

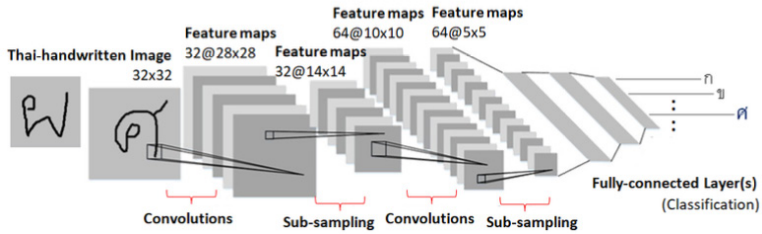

Fig.5. ConvNet architecture (an input layer, convolution, and max-pooling layers, and fully-connected layer)

4.1. Convolution layer

The feature maps of the previous layer are convolved with the kernel (a.k.a. weight, filter, or mask) such as Gaussian or Gabor. Moreover, the output is sent out to the activation functions (Mookdarsanit & Mookdarsanit, 2019b). To formulate the output feature maps by

$$x_j^l = \left(\sum_{i \in M_j} x_j^{l-1} k_j^l + b_j^l \right), \tag{21}$$

where x_j^l and x_j^{l-1} are the output of current and previous layer, k_j^l is the j-th kernel in the present layer, b_j^l is j-th bias value in the present layer, M_j is a selection of input maps.

4.2. Sub-sampling layer

The sub-sampling executes the down-sampling operation ($downsamp(\bullet)$) over the input maps. The number of output maps equals that of the input map, but the size is reduced according to the down-sampling mask

$$x_j^l = f\left(\beta_j^l downsamp(x_j^{l-1}) + b_j^l\right). \tag{22}$$

4.3. Classification layer

Since the masks are updated during the convolutional operation between the convolutional layer and the previous layer on the feature maps, to keep with this, the weights of each layer are also computed. This layer computes the probability score for each class of unknown objects using the convolutional layer's extracted features. Furthermore, the classification layers still have a gap to reduce the network complexity by its fully-connected structure proposed in section 4.4.

4.4. Our modified version

Since the flat ConvNet for Thai-handwritten recognition is too expensive in computation and time processing in the larger network. In this paper, we propose a ConvNet with dropout named “ThaiWrittenNet” to reduce the model complexity with the help of deep belief network (DBN) by reconstructing and adapting the parameters (a.k.a. aggregation) in fully-connected classification layers (Hinton, Osindero & Teh, 2006), as shown in Figure 6.

The weight change (Δw_{ij}) can be computed by learning rate function ($\varepsilon(\bullet)$) with parameters: observable variables (v_i) and hidden variables (h_i) from ConvNet either ConvNet with DBN, as mathematically described by

$$\Delta w_{ij} = \varepsilon\left(\langle v_i, h_i \rangle_{fully-connected}, \langle v_i, h_i \rangle_{\mathrm{mod}\,ified}\right). \tag{23}$$

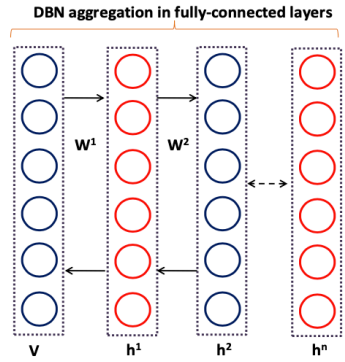

Fig. 6. DBN aggregation of Thai handwritten recognition by ConvNet

5. Experimental results and discussion

For the experimental results, we categorize into 3 groups. The dataset ThaiWrittenNet as our primary data contains 9,282 Thai-handwritten images that consist of 87 classes: 44 consonants, 18 vowels, 5 diacritics, 4 tone marks, 10 Thai numerals, and 6 special symbols. The comparison is made under 7,426 image training and 1,856 image testing.

5.1. Comparison between matching under k-NN

This part only compares many SIFT versions, SURF, and HoG under k-NN lazy learning, where k=3, 4, and 5. The results are shown in Table 1.

Table 1: Accuracy comparison of our dataset under k-NN

Feature Extraction	Accuracy k-NN	
	k=5	k=10
128-D SIFT	87.89	85.27
64-D SIFT	80.56	81.05
32-D SIFT	68.85	70.78
SURF	78.41	78.58
HoG (or Dense-SIFT)	78.67	82.02

Most feature extraction algorithms work better if k=10, except for 128-D SIFT. Since k-NN is lazy learning without any training function, the k number helps to rank similar images from the vector. Although SURF has 64 dims as well as 64-D SIFT, SIFT still performs in higher accuracy. 128-D SIFT has the highest accuracy for k-NN as its highest dimensions with the most complexity. SURFs in both k are not so different values. HoG should have worked with some machine learning like SVM.

5.2. Comparison of traditional machine learning

Those handcraft features for Thai-handwritten recognition/recognition, e.g., SIFT, SURF, and HoG, are compared under the Non-lazy machine learning like, e.g., MLP and SVM, as shown in Table 2.

Most Thai-handwritten recognition algorithms are suitable for SVM, rather than MLP. Since SVM is well-performed on high dimensionality. Although HoG is just an image description without detection, it works well together with SVM. Owing to the working well in the high dimensional vector of SVM, the combined handcrafted model like HoG with 64-D SIFT, HoG with 32-D SIFT, and HoG with SURF can improve the recognition accuracy but they are more complexity and computational resource.

Table 2: Accuracy comparison of our dataset under different machine learning models

Feature Extraction	Accuracy Traditional Machine Learning	
	MLP	SVM
64-D SIFT	83.28	86.42
32-D SIFT	83.70	82.28
SURF	81.66	87.93
HoG (or Dense-SIFT)	80.76	91.78
HoG with 64-D SIFT	88.87	95.16
HoG with 32-D SIFT	86.14	93.51
HoG with SURF	89.39	96.83

5.3. Overall experimental results

The configuration of parameters (layer operation, number of feature maps, size of feature maps, size of windows, and number of parameters) in our ConvNet as ThaiWrittenNet is shown in Table 3.

Table 3: Our ConvNet configuration as ThaiWrittenNet

Layer	Operation	No. of Feature Maps	Size of Feature Maps	Size of window	No. of Parameters
C1	Convolution	32	28x28	5x5	832
S1	Max-pooling	32	14x14	2x2	-
C2	Convolution	64	10x10	5x5	53,248
S2	Max-pooling	64	5x5	2x2	-
F1	Fully-connected	312	1x1	-	519,168
F2	Fully-connected	10	1x1	-	3,130

Deep learning and traditional machine learning are compared in Table 4. It is obviously seen that deep learning methods provide better results in accuracy than traditional machine learning with handcrafted feature extraction. Since Thai-

handwritten images have only the feature representation of written-line over the background, Gabor filter is more suitable for these handwritten images than Gaussian.

Table 4: Overall comparison

Algorithm	Accuracy
128-D SIFT + kNN (k=5)	87.89
32-D SIFT + MLP	83.70
HoG + SVM	91.78
HoG with 64-D SIFT + SVM	95.16
HoG with 32-D SIFT + SVM	93.51
HoG with SURF + SVM	96.83
Gaussian + ConvNet	96.57
Gabor + ConvNet	97.01
Gaussian + ThaiWrittenNet	98.18
Gabor +ThaiWrittenNet	98.59

6. Conclusion

In this paper, we propose a novel ThaiWrittenNet based on Convolutional Neural Network (ConvNet or CNN) with Deep Belief Network (DBN) for Thai handwritten recognition. We also compare our ThaiWrittenNet to traditional machine learning with handcrafted feature extraction by our primary dataset. The dataset contains 9,282 Thai-handwritten images that consist of 87 classes: 44 consonants, 18 vowels, 5 diacritics, 4 tone marks, 10 Thai numerals, and 6 special symbols. The comparison is made under 7,426 image training and 1,856 image testing. From the experimental results, ConvNet based outperformed those traditional machine learning.

Moreover, DBN can be used to reduce network complexity and provide higher accuracy. For future work, deep learning can share parameters in learning tasks at different times, known as domain adaptation – that is an efficient learning model to learn a variety of Thai handwritten styles. Moreover, many augmentation and generative models can be used to generate more Thai-handwritten images to enhance accuracy.

Acknowledgment

The paper “ThaiWrittenNet: Thai Handwritten Script Recognition using Deep Neural Networks” was proposed to integrate Thai language and computer vision as a Thai linguistic heritage application. All local Thai handwritten images in this paper were watermarked and copyrighted as our primary data. The reader(s) can request our local collection via the email (in TERM OF USE). The hardware and other resources were dedicated to Chandrakasem Rajabhat University, Bangkok, Thailand.

References

Ager, S. (2020). How many languages are there in the world?. Retrieved from: https://www.ethnologue.com/guides/how-many-languages

Alom, Md. Z., Sidike, P., Taha, T. M. & Asari, V. K. (2017). Handwritten Bangla digit recognition using deep learning, *arXiv:1705.02680.*

Alom, Md. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Esesn, B. C. V., Awwal, A. A. S. & Asari, V. K. (2018). The history began from AlexNet: A comprehensive survey on deep learning approaches. *arXiv:1803.01164.*

Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404-417). Springer, Berlin, Heidelberg.

Boonkwan, P., & Supnithi, T. (2017, June). Bidirectional deep learning of context representation for joint word segmentation and POS tagging. In *International Conference on Computer Science, Applied Mathematics and Applications* (pp. 184-196). Springer, Cham.

Chaiwatanaphan, S., Pluempitiwiriyawej, C., & Wangsiripitak, S. (2017). Printed Thai character recognition using shape classification in video sequence along a line. *Engineering Journal, **21***(6), 37-45.

Cortes, C., & Vapnik, V. (1995). *Support-vector networks. Machine learning, 20*(3), 273-297.

Daengsi, T., & Wuttidittachotti, P. (2019). QoE Modeling for Voice over IP: Simplified E-model Enhancement Utilizing the Subjective MOS Prediction Model: A Case of G. 729 and Thai Users. *Journal of Network and Systems Management, 27*(4), 837-859.

Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

Emsawas, T., & Kijsirikul, B. (2016, August). Thai printed character recognition using long short-term memory and vertical component shifting. In *Pacific Rim International Conference on Artificial Intelligence* (pp. 106-115). Springer, Cham.

Fukushima, K. (1980). Biological cybernetics neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. ***Biol Cybern*, *36*,** 193-202.

Haruechaiyasak, C., Kongthon, A., Palingoon, P., & Trakultaweekoon, K. (2013, October). S-Sense: A sentiment analysis framework for social media sensing. In *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP)* (pp. 6-13).

Haruechaiyasak, C., Kongyoung, S., & Dailey, M. (2008, May). A comparative study on Thai word segmentation approaches. In *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (Vol. 1, pp. 125-128). IEEE.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation, 18*(7), 1527-1554.

Inthajakra, L., Prachyapruit, A., & Chantavanich, S. (2016). The Emergence of Communication Intellectual History in Sukhothai and Ayutthaya Kingdom of Thailand. *Social Science Asia, 2*(4), 32-41.

Ismayilov, E. & Mammadov, R. (2019). Parallel solution of features subset selection process for hand-printed character recognition. *Azerbaijan Journal of High Performance Computing, 2*(2), 170-177.

Klahan, A., Pannoi, S., Uewichitrapochana, P., & Wiangsripanawan, R. (2018, July). Thai Word Safe Segmentation with Bounding Extension for Data Indexing in Search Engine. In *International Conference on Computing and Information Technology* (pp. 83-92). Springer, Cham.

Koanantakool, H. T., Karoonboonyanan, T., & Wutiwiwatchai, C. (2009). Computers and the thai language. *IEEE Annals of the History of Computing, 31*(1), 46-61.

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).

Lapjaturapit, T., Viriyayudhakom, K., & Theeramunkong, T. (2018, May). Multi-candidate word segmentation using bi-directional LSTM neural networks. In *2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES)* (pp. 1-6). IEEE.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324.

Liu, D., Bober, M. & Kittlet, J. (2019). Visual semantic information pursuit: a survey. *arXiv:1903.05434*

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep learning for generic object detection: A survey. ***International journal of computer vision***, ***128***(2), 261-318.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision, 60*(2), 91-110.

Lyons, S. (2016, August). Quality of Thai to English Machine Translation. In Pacific Rim Knowledge Acquisition Workshop (pp. 261-270). Springer, Cham.

Mookdarsanit, L. & Mookdarsanit, P. (2019a). SiamFishNet: The deep investigation of Siamese fighting fishes. *International Journal of Applied Computer Technology and Information Systems, 8*(2), 40-46.

Mookdarsanit, L. & Mookdarsanit, P. (2019b). Thai herb recognition with medicinal properties using convolutional neural network. *Suan Sunandha Science and Technology Journal, 6*(2), 34-40.

Mookdarsanit, L. & Mookdarsanit, P. (2020a). An adversarial perturbation technique against reCaptcha image attacks. *Journal of Science and Technology Buriram Rajabhat University (on print).*

Mookdarsanit, L. & Mookdarsanit, P. (2020b). The insights in computer literacy to-

ward HR intelligence: Some associative patterns between IT subjects and job positions. *Journal of Science and Technology RMUTSB (on print).*

Mookdarsanit, L. (2020). The intelligent genuine validation beyond online Buddhist amulet market. *International Journal of Applied Computer Technology and Information Systems, 9*(2), 7-11.

Mookdarsanit, P. & Mookdarsanit, L. (2018a). A content-based image retrieval of Muay-Thai folklores by salient region matching. *International Journal of Applied Computer Technology and Information Systems, 7*(2), 21-26.

Mookdarsanit, P. & Mookdarsanit, L. (2020). The autonomous nutrient and calorie analytics from a Thai food image. *Journal of Faculty Home Economics Technology RMUTP (on print).*

Mookdarsanit, P. & Rattanasiriwongwut, M. (2017b). Location Estimation of a Photo: A Geo-signature MapReduce Workflow. *Engineering Journal, 21*(3), 295-308.

Mookdarsanit, P. & Rattanasiriwongwut, M. (2017c). MONTEAN Framework: a magnificent outstanding native-Thai and ecclesiastical art network. *International Journal of Applied Computer Technology and Information Systems, 6*(2), 17-22.

Mookdarsanit, P. (2019). TGF-GRU: A Cyber-bullying Autonomous Detector of Lexical Thai across Social Media. NKRAFA JOURNAL OF SCIENCE AND TECHNOLOGY, 15, 50-58.

Mookdarsanit, P., & Gertphol, S. (2013, January). Light-weight operation of a failover system for Cloud computing. In *2013 5th International Conference on Knowledge and Smart Technology (KST)* (pp. 42-46). IEEE.

Mookdarsanit, P., & Ketcham, M. (2016, February). Image Location Estimation of well-known Places from Multi-source based Information. In *The 11th International Symposium on Natural Language Processing* (p. 75).

Mookdarsanit, P., & Mookdarsanit, L. (2018). Contextual Image Classification towards Metadata Annotation of Thai-tourist Attractions. *ITMSoc Transactions on Information Technology Management, 3*(1), 32-40.

Mookdarsanit, P., & Mookdarsanit, L. (2018). Name and recipe estimation of thai-desserts beyond image tagging. *Kasem Bundit Engineering Journal, 8,* 193-203.

Mookdarsanit, P., & Mookdarsanit, L. (2018b). An Automatic Image Tagging of Thai Dance's Gestures. In *Joint Conference on ACTIS & NCOBA, Ayutthaya, Thailand* (pp. 76-80).

Mookdarsanit, P., & Rattanasiriwongwut, M. (2017, January). GPS Determination of Thai-temple Arts from a Single Photo. In *The 11th International Conference on on Applied Computer Technology and Information Systems, Bangkok, Thailand* (pp. 42-47).

Mookdarsanit, P., Soimart, L., Ketcham, M., & Hnoohom, N. (2015, November). Detecting image forgery using XOR and determinant of pixels for image forensics. In *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 613-616). IEEE.

Nomponkrang, T., & Sanrach, C. (2016). The Comparison of Algorithms for Thai-Sen-

tence Classification. *International Journal of Information and Education Technology, 6*(10), 801-808.

Olaode, A. & Naghdy, G. (2020). Adaptive bag-of-visual word modelling using stacked-autoencoder and particle swarm optimisation for the unsupervised categorisation of images. *IET Image Processing.* doi:10.1049/iet-ipr.2019.1160

Olaode, A., Naghdy, G. & Todd, C. (2014). Unsupervised classification of images: A review. *International Journal of Image Processing, 8*(5), 325-342.

Pornpanomchai, C., Wongsawangtham, V., Jeungudomporn, S. & Chatsumpun, N. (2011). Thai handwritten character recognition by genetic algorithm. *IACSIT International Journal of Engineering and Technology, 3*(2), 148-153.

Raghu, M. & Schmidt, E. (2020). A survey of deep learning for scientific discovery, *arXiv:2003.11755.*

Rathi, R., Pandey, R. K., Chaturvedi, V. & Jangid, M. (2012). Offline handwritten Devanagari vowels recognition using KNN classifier. *International Journal of Computer Applications, 49*(23), 11-16.

Rumelhart, D. E. & McClelland, J. L. (1987). Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations.* (pp. 318-362).

Satienkoses, Y. (1981). Essays on Thai folklore. Duang Kamol, Bangkok.

Soimart, L., & Ketcham, M. (2015, January). The segmentation of satellite image using transport mean-shift algorithm. In *13th International Conference on IT Applications and Management (ITAM-13)* (pp. 124-128).

Soimart, L. & Ketcham, M. (2016a). An efficient algorithm for earth surface interpretation from satellite imagery. *Engineering Journal, 20*(5), 215-228.

Soimart, L., & Ketcham, M. (2016, February). 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* (p. 74).

Soimart, L. & Mookdarsanit, P. (2016a). Gender estimation of a portrait: Asian facial-significance framework. In *Proceedings of the 6th International Conference on Sciences and Social Sciences.* Mahasarakham, Thailand.

Soimart, L., & Mookdarsanit, P. (2016). Multi-factor authentication protocol for information accessibility in flash drive. *The 9th Applied Computer Technology and Information Systems, Nakhon Pathom,* 10-13.

Soimart, L. & Mookdarsanit, P. (2017a). Ingredients estimation and recommendation of Thai-foods. *SNRU Journal of Science and Technology, 9*(2), 509-520.

Soimart, L. & Mookdarsanit, P. (2017b). Name with GPS auto-tagging of Thai-tourist attractions from an image. In *Proceedings of the 2nd Technology Innovation Management and Engineering Science International Conference* (pp. 211-217). Nakhon Pathom, Thailand.

Soimart, L. & Pongcharoen, P. (2011). Multi-row machine layout design using aritificial bee colony. In *Proceedings of 2011 International Proceedings of Economics Develop-*

ment & Research. Bangkok, Thailand: IPEDR

Sornlertlamvanich, V., Potipiti, P, Wutiwiwatchai, C. & Mittrapiyanuruk, P. (2000). The state of the art in Thai language processing. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics* (pp. 1-2). Stroudsburg, PA, USA: ACM.

Srihari, S. N. & Kuebert, E. J. (1997). Integration of hand-written address interpretation technology into the United States Postal Service Remote Computer Reader system. In *Proceedings of the 4th International Conference on Document Analysis and Recognition* (pp. 892-896). ACM.

Surinta, O. & Nitsuwat, S. (2006). Handwritten Thai character recognition using Fourier descriptors and robust C-prototype. *Information Technology Journal. 2*(1), 92-96

Surinta, O., Karaaba, M. F., Schomaker, L. R. B. & Wiering, M. A. (2015). Recognition of handwritten characters using local gradient feature descriptors. *Engineering Applications of Artificial Intelligence, 45,* 495-414.

Theeramunkong, T., & Tanhermhong, T. (2004). Pattern-based features vs. statistical-based features in decision trees for word segmentation. *IEICE TRANSACTIONS on Information and Systems, 87*(5), 1254-1260.

Torfi, A., Shivani, R. A., Keneshloo, Y., Tavvaf, N. & Fox, E. A. (2020). Natural language processing advancements by deep learning: A survey, *arXiv:2003.01200.*

Wat Chonprathan Rangsarit. (2001). A Buddhism preachment by Phra Phrom Mangkhalachan. *Department of Science Service Journal. 49*(155), 33-34. (in Thai).

World Bank. (2018). Population, Total. Retrieved from: https://data.worldbank.org/indicator/SP.POP.TOTL?locations=TH

Zheng, L., Yang, Y. & Tian, Q. (2017). SIFT meets CNN: A decade survey of instance retrieval. *arXiv:1608.01807.*

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object detection in 20 years: A survey. *arXiv:1905.05055.*

Submitted 23.03.2020
Accepted 23.05.2020