**International Journal of Computing and Digital Systems**
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (May-2021)

http://dx.doi.org/10.12785/ijcds/100165

# ThaiWritableGAN: Handwriting Generation under Given Information

**Lawankorn Mookdarsanit[1*] and Pakpoom Mookdarsanit[2]**

[1*] *Faculty of Management Science, Chandrakasem Rajabhat University, Bangkok, Thailand*
[2] *Faculty of Science, Chandrakasem Rajabhat University, Bangkok, Thailand*

*Received 08 Dec.2020, Revised 23 Mar. 2021, Accepted 03 Apr. 2021, Published 02 May. 2021*

**Abstract:** For the local technique challenge, Thai has different symbols' vertical positions with no space between characters and words. Thai handwriting recognition has been a long time research problem. To join the edge between unsupervised Generative adversarial network (GAN) and Thai handwriting recognition, this paper introduces a novel Thai handwriting generation under given information (named "ThaiWritableGAN"). ThaiWritableGAN is proposed to map textual information with real handwritten data to generate a new handwritten style (a.k.a. calligraphy). The proposed algorithm consists of generator (G), discriminator (D), and recognizer (R). The synthesized (or generated) handwritten sample is done by G which is proposed to fool D. D is assigned to discriminate an unknown handwritten image that it is real or generated. R is a convolutional neural network (pre-trained by real Thai handwritten images) that is additionally added to recognize the synthesized images (generated by G). For the scientific knowledge discovery, self-attention mechanism (introduced by Google AI) and R totally boost more realistic Thai handwriting generation as well as other languages. The gradient balancing argument should be set to 1. The word error rate (WER) can be relieved by computational reduction in R's gradient. But the reduction affects a little lower realistic quality of Thai handwriting, measured by Fréchet inception distance (FID). For the beyond following, Thai handwriting generation competition might be opened that the local researchers can submit their handwriting generation algorithms to the calligraphy contest.

**Keywords:** Thai Handwriting Generation, Conditional Adversarial Network, Handwritten Text Synthesis, Generative Model

## 1. INTRODUCTION

From the historical evidence, Thai scripts were firstly composed on Sukothai stone [1] inscriptions by King Ramkhamhaeng [2] longer than 720 years ago. Thai scripts had 44 consonants, 18 vowels, 5 diacritics, 4 tone marks, 10 Thai numerals, and 6 special symbols [3]. During Rattanakosin period, many Thai literatures and quotes [4] were composed by Sunthorn Phu – known as Thailand's Shakespeare [5] who was one of the world great poets [4-5] by UNESCO in 1986. Even now, these quotes were included in many textbooks for teaching all Thai students as their literacy [6]. Thai text had no space between "*the (i-1)-th word*", "*the (i)-th word*" and "*the (i+1)-th word*" as a challenge in word-level embedding. For the character-level challenge, all symbols had many vertical positions. What's more, one Thai word had 5 tonal pronunciations with different meanings.

For the linkage between Thai language and computer vision, both Thai printed text and handwriting recognition was a long time interesting area in Thai natural language processing (Thai-NLP).

Thai printed text recognition [7-8] was normally included in many scanned-document applications [8] that it had no more challenge. In the same way, Thai handwriting recognition [9] was also one of Thai-NLP's hot topics longer than 20 years ago. Frankly, Convolutional neural network (CNN) showed the handwriting recognition accuracy higher than 95% [10]. Both handwriting and printed text recognition was such an old problem in Thai-NLP. Especially in the deep learning era, big data in our daily life was not only used for supervised image recognition but also unsupervised generation. The popular generative model was Generative adversarial network (GAN) [11] that enabled the synthesis of diverse images.

To join the edge between unsupervised GAN and Thai handwriting recognition, this paper originally introduced a novel generative model (named "ThaiWritableGAN") to generate Thai handwritten text under given information based on Geometric GAN [12].

*E-mail: lawankorn.s@chandra.ac.th, pakpoom.m@chandra.ac.th*

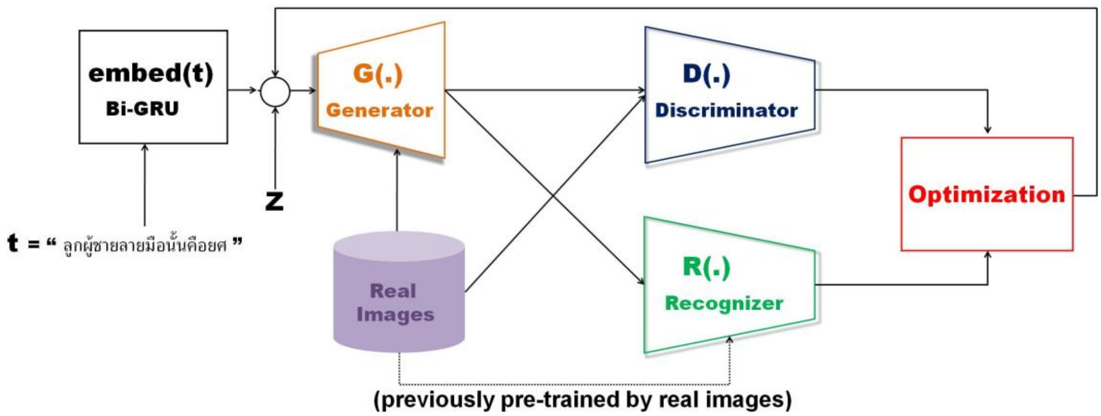


Figure 1. Proposed ThaiWritableGAN workflow based on unsupervised GAN: the model consists of generator (G), discriminator (D) and recognizer (R), respectively

Based on the competition between generator (G) and discriminator (D) in GAN, the recognizer (R) was additionally added to recognize Thai textual image as the realistic and readable quality. The major contribution could be summarized as:

- The realistic Thai handwriting generation could be reformed by self-attention mechanism and additional recognizer.
- The best gradient balancing value between R and D for Thai handwriting generation was 1.
- For the proposed Thai handwriting generation, the computational reduction in R’s gradient could relieve the word error rate (WER) but little affect the realistic quality.

This paper was organized as follows. Related works were in part 2. Part 3 talked about the proposed ThaiWritableGAN. Experimental settings and results were in part 4. And part 5 was conclusion.

## 2. Related Works

Both Thai printed text [12] and handwriting recognition [13-14] referred to a conversion of physical textual images into textual information [8]. The related works could be divided into 2 groups: Thai printed text recognition and From Thai handwriting recognition to generation, respectively.

### A. *Thai printed text recognition*

A technical document (written by non-native Thai author) from Thailand's National Electronics and Computer Technology Center (NECTEC) firstly introduced “Thai font recognition” [7] as a new computer science problem in 1996. Meantime, the information composed by manually physical handwriting was gradually outdated that it was the beginning of digital era. Initially, NECTEC researchers demonstrated the performance of Multi-layer perceptron (MLP) for state-of-the-art Thai printed text recognition [8] with the Thai printed text software project [15]. The part of speech (POS), tri-gram model and Winnow algorithm [16] were basically combined to improve the quality of Thai printed text recognition. Coupled with Kohonen self-organization and back-propagation in two-step neural network [17] was proposed to recognize Thai and English printed texts. However, the processing resource during those days was too expensive to implement those neural architectures [8, 15-17] as supervised learning. Some Thai printed text recognition algorithms were based on handcrafted feature matching with other supervised models such as Euclidean-based lazy learning [18] and Fuzzy with rough set [19]. The rough set was essential for Thai textual invariant features [20]. One paper introduced the combination of back propagation and Fuzzy logic [21] that could be enhanced by boundary normalization [22] as new ensemble learning for Thai printed text recognition. With the help of dimension reduction, Principal component analysis (PCA) [23] was used to select only the useful features for matching with unknown printed texts. With the well-working on a large number of features, Support vector machine (SVM) – a kernel based classification that demonstrated the acceptable accuracy [24]. As well as “ImageNet Large Scale Visual Recognition Challenge (ILSVRC)”, there also was a local algorithm competition in Thailand, known as “Benchmark for Enhancing the Standard of Thai Language Processing (BEST)” organized by NECTEC. As referred to the “BEST 2013” [25] with the challenge on Thai printed text recognition, the winners in the set #2 and #3 (from Chiang Mai University) applied the kernel based algorithm that produced the highest accuracy, totally higher than MLP and lazy learning. Unlike the diverse styles of Thai handwriting from different writers, the printed text recognition was only such the static shapes in different fonts. Handwriting was absolutely harder than static printed texts. It was not surprise that Thai printed text recognition with high correctness was already available in a way of scanned-document applications with other optional language functions [26] or in VDO sequence [27]. For this reason, the last BEST 2019 project was emphasized on Thai handwriting recognition [28].

### *B. From Thai handwriting recognition to generation*

Handwriting recognition could be seen as a subclass of sketch recognition. Free-hand sketches easily caused a great variability of written styles by different writers; especially in Thai scripts with different positions of vowels, diacritics, and tone marks. Thai handwriting recognition [13-14] was a hot area in Thai natural language processing (Thai-NLP) [9]. Early, Fuzzy logic was used in many researches [29-30] to identify the variant Thai handwritten features. As well as Multi-layer perceptron (MLP), Thai handwritten feature was extracted by Fourier descriptor, learned by MLP; together [31] with optimized by Genetic algorithm (GA) [32], and the dimension reduced by Principal component analysis (PCA) [33]. The unsupervised clustering of geometric similarity was used to pre-process [34] before MLP learning. By combination, MLP and Fuzzy operations (a.k.a. Neuro-fuzzy) [35] were also used for learning and approximating the uncertain Thai handwritten feature, respectively. A variety of geometric feature methods [36-39] were matched and then the similar patterns were ranked. For the crucial comparison using the same handcrafted feature extraction, MLP was shown to be higher accuracy than Euclidean-based lazy learning [40]. Not only MLP but also Support vector machine (SVM) provided good results with Thai handwriting recognition [41]. Nevertheless, MLP and SVM were defeated by deep learning method, especially in large-scale data. The beginning of deep learning referred to "AlexNet" in 2012 that demonstrated the Convolutional neural network (CNN) [42] better than traditional handcrafted feature extraction methods, especially in large-scale image dataset. CNN was said to the world as a deep learning for computer vision. As a matter of fact, the first version of CNN (called "LeNet") was used to recognize handwritten digits/letters with higher accuracy than 95% [10] in 1998. LeNet was not so much popular because of the limited processing resource. For Thai handwriting recognition [43], CNN with different architectures (e.g., VGGNet, Inception and ResNet) was also proved to be higher accuracy than handcrafted feature (e.g., SIFT and HoG) with SVM. As the higher accuracy than 95%, Thai handwriting recognition seemed to be not anymore harder since the rebirth of CNN in 2012 [42]. In 2019, CNN with the sequence of Thai letters in Recurrent Neural Network (RNN) could be combined and trained using Connectionist Temporal Classification (CTC) loss [44]. Furthermore, some types of RNN: Long-short-term memory (LSTM) [45] and Gated Recurrent Unit (GRU) [3] had already shown the friendliness and usefulness of Thai word understanding. As referred to the BEST 2019 competition by NECTEC, the 4/5 of candidates applied GRU to different CNN architectures that demonstrated the compatibility for Thai handwritten recognition [27].

For the first concept of data generation, the change of computer vision society occurred again when unsupervised Generative adversarial networks (GAN) [11] enabled the data generation. Prior to GAN, a sequence to sequence supervised learning like LSTM had been used to generate English word but it was not compatible for handwritten features like stroke-written variability [46]. A type of conditional GAN for image-to-image translation (the well-known name as "pix2pix") [47] was used to generate many images in 2017. Pix2pix's concept was to map between input and output images by similar edge strokes. From handwriting recognition to generation, the pix2pix based concept was initially applied to Chinese calligraphy generation [48], known as "zi2zi". Later, Chinese handwritten characters [49] could be generated by "DenseNet-5 CycleGAN". Such a given English textual content also could be composed [50] into calligraphies (in different writing styles) [51] as well as French and Arabic [52]. Obviously, calligraphy/handwriting generation has been interesting since the GAN was introduced to the world.

Above all, it is time to firstly introduce a novel algorithm to generate Thai handwriting under given information as a new research topic in Thai language with computer vision.

## 3. PROPOSED THAIWRITABLEGAN

According to the basic concept of original Generative adversarial network (GAN) [11], it is the competition between generator (G) and discriminator (D). G tries to map some noise (z) into the real handwriting image (as handwriting or calligraphy synthesis) that is proposed to fool the D. In contrast, D is assigned to discriminate the handwritten image that is real or generated. Additionally, the recognizer (R) [50] is previously pre-trained by real handwriting images; R is used to recognize the synthesized Thai handwriting images (generated by G). For training ThaiWritableGAN, the adversary is G and D in the training competition. The finished completion of the trained model is that D cannot anymore discriminate the difference between generated and real handwriting (when G totally conquers D); and R is also able to recognize the generated handwriting correctly. The proposed algorithm is shown in Fig. 1.

### *A. Generator*

Since each Thai script is influenced by its predecessor and/or successor in vertical and horizontal positions. Bi-directional gated recurrent unit (Bi-GRU) is used for Thai character-level embedding into a sequence of characters.

Following the GAN paradigm, the sequence of characters with a pile of noise is concatenated and sent to all residual blocks (ResBlocks) within the generator

(G). G consists of many up-sampling ResBlocks with self-attention mechanism, as shown in Fig. 2

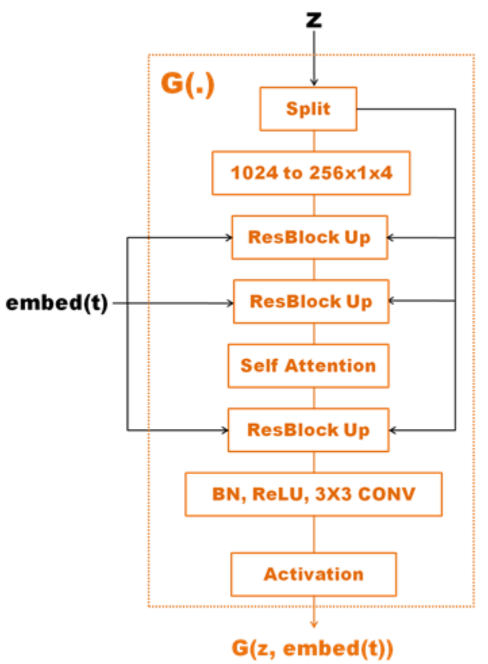


Figure 2. Architecture of generator (G)

To generate 512x128 Thai handwritten images to discriminator (D) and recognizer (R), the activation function in G is in a normal hyperbolic tangent form. Technically, self-attention provides the high score of correlation [53] and shows the great success in the mechanism of Transformer (by Google AI [54]).

For the proposed algorithm, bi-linear interpolation [55] is used as an up-sampling function. The loss function in G ( $L_{(G,embed(t))}$ ) can be computed by (1).

$$L_{(G,embed(t))} = -\mathrm{E}_{z\sim p_z, t\sim p_w}\left[D(G(z, embed(t)))\right] + \mathrm{E}_{z\sim p_z, t\sim p_w}\left[CTC(t, R(G(z, embed(t))))\right] \quad (1)$$

where $\mathrm{E}_{z\sim p_z, t\sim p_w}$ is expected value over noise ( $z$ ) and a given Thai text ( $t$ ), $D(\bullet)$ as discriminator, $G(\bullet)$ as generator, $embed(t)$ as a 4-layer Bi-GRU to embedding a given Thai textual content and $CTC(\bullet)$ as CTC loss function

*B. Discriminator*

Adversarially, the discriminator (D) [51] is designed to estimate the probability that Thai handwritten image is real or generated by generator (G). D consists of many down-sampling blocks by average pooling [52] and also based on self-attention like G, as Fig. 3. The loss in D ( $L_D$ ) can be computed by (2).

$$L_D = -\mathrm{E}_{(x,t)\sim p_{data}}\left[\min(0, -1 + D(x))\right] - \mathrm{E}_{z\sim p_z, t\sim p_w}\left[\min(0, -1 - D(G, z(embed(t))))\right] \quad (2)$$

where $\mathrm{E}_{(x,t)\sim p_{data}}$ is expected value over real image ( $x$ ) and a given Thai text ( $t$ ), $\mathrm{E}_{z\sim p_z, t\sim p_w}$ is expected value over noise ( $z$ ) and a given Thai text ( $t$ ), $D(\bullet)$ as discriminator, $G(\bullet)$ as generator, $embed(t)$ as a 4-layer Bi-GRU to embedding a given Thai textual content

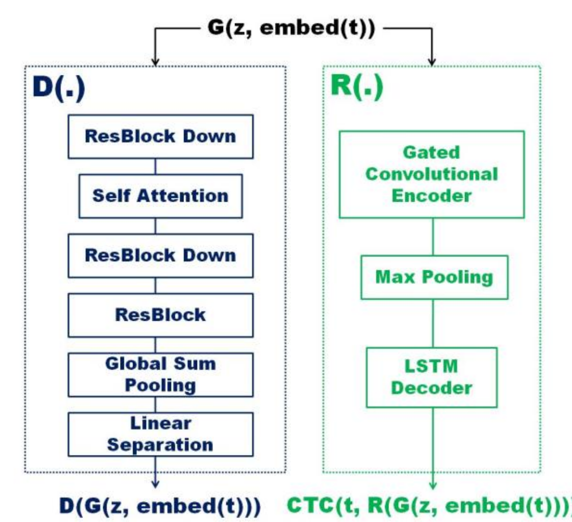


Figure 3. Architecture of discriminator (D) and recognizer (R)

*C. Recognizer*

Original GAN has no recognizer. The recognizer (R) implies [50] the realistic and readable textual image. The recognized Thai text from R is usefully parameterized as penalty in generator.

**Algorithm 1** ThaiWritableGAN: handwriting generation under given information

**Input:** $t \leftarrow$ a given Thai textual information, $z \leftarrow$ noise , $x \leftarrow$ real image

**Output:** The generator parameterized setting to generate Thai handwriting under a given content as $x_{generated}$

1: Call $embed(t)$; ▷ embed a given Thai textual content by 4-layer Bi-GRU
2: **repeat**
3: **if** $\left((\nabla D \wedge \nabla R) \notin \phi\right)$ **then**
4: $L_{(G,embed(t))} \leftarrow -\mathrm{E}_{z\sim p_z, t\sim p_w}\left[D(G(z,embed(t)))\right] + \mathrm{E}_{z\sim p_z, t\sim p_w}\left[CTC(t, R(G(z,embed(t))))\right]$;
5: $z \sim \mathrm{N}(0,1)$; ▷ an independent and identically distributed (iid) Gaussian normal matrix
6: **end if**
7: $x_{generated} \leftarrow G(z,embed(t))$; ▷ generated Thai handwritten samples as $x_{generated}$
8: Call $D(x)$;
9: Call $D(x_{generated})$;
10: Call $R(x_{generated})$;
11: $\nabla D \leftarrow -\frac{\partial D(x_{generated})}{\partial x_{generated}}$;
12: $\nabla R \leftarrow \frac{\partial CTC(s, R(x_{generated}))}{\partial x_{generated}}$;
13: $\nabla R \leftarrow \alpha\left(\frac{\sigma(\nabla D)}{\sigma(\nabla R)}(\nabla R - \mu(R)) + \mu(D)\right) \vee \alpha\left(\frac{\sigma(\nabla D)}{\sigma(\nabla R)} \bullet \nabla R\right)$; ▷ $\nabla R$ computed by Eq.4 or 5
14: **until** the acceptance criteria $\left(D(G(z,embed(t))) \geq 0.5\right) | \left(CTC(t, R(G(z,embed(t)))) \geq 0.8\right)$

As well as [52], the R is encoded by Gated convolutional neural network (Gated-CNN) with 5 convolutional layers and decoded by 2 layers of stacked bidirectional Long-short-term memory (stacked Bi-LSTM). The $L_R$ is previously pre-trained by real Thai handwritten images based on Connectionist temporal classification (CTC) function, as (3).

$$L_R = -\mathrm{E}_{(x,t)\sim p_{data}}\left[CTC(t, R(x))\right] \tag{3}$$

where $\mathrm{E}_{(x,t)\sim p_{data}}$ is expected value over real image ( $x$ ) and a given Thai text ( $t$ ), $R(x)$ as recognizer and $CTC(\bullet)$ as CTC loss function

*D. Optimization*

The generator ( $G(z,embed(t))$ ) is optimized to generate Thai handwriting images by the loss of discriminator (D) and recognizer (R) [50], as Algorithm 1 (named "ThaiWritableGAN"). The model is trained around hundred thousand iterations with the gradient clipping in D and R. Since the gradient stemming from $\|\nabla R\|$ is 100-1000 times greater than $\|\nabla D\|$ [52], the standard deviation and average of discriminator loss and recognizer loss is applied to the new gradient of $\nabla R$ for the balance between D and R. The different $\nabla R$ 's gradient formulas may affect the algorithm's results. The full version of $\nabla R$ can be computed by (4).

$$\nabla R \leftarrow \alpha\left(\frac{\sigma(\nabla D)}{\sigma(\nabla R)}(\nabla R - \mu(\nabla R)) + \mu(\nabla D)\right) \tag{4}$$

where $\mu(\bullet)$ and $\sigma(\bullet)$ refers to average and standard deviation, $\alpha$ as gradient balancing between $D$ and $R$ , the $\nabla D$ gradient as

$\nabla D = -\dfrac{\partial D(x_{generated})}{\partial x_{generated}}$ and the former $\nabla R$ gradient as $\nabla R = \dfrac{\partial CTC(s, R(x_{generated}))}{\partial x_{generated}}$

For the computational reduction, only the standard deviation has high variability enough for the optimization, the average terms in both D and R might be eliminated into (5).

$$\nabla R \leftarrow \alpha\left(\frac{\sigma(\nabla D)}{\sigma(\nabla R)} \bullet \nabla R\right) \tag{5}$$

## 4. Experimental Settings and Results

Based on the joint edge between Generative adversarial network (GAN) and Thai handwriting recognition, this part describes all about the experiments and discussions.

| | | |
|---|---|---|
| Consonants | 44 | ก ข ฃ ค ฅ ฆ ง<br>จ ฉ ช ซ ฌ ญ<br>ฎ ฏ ฐ ฑ ฒ ณ<br>ด ต ถ ท ธ น<br>บ ป ผ ฝ พ ฟ ภ ม<br>ย ร ล ว ศ ษ ส<br>ห ฬ อ ฮ |
| Vowels | 18 | ะ ◌ั า ◌ำ ◌ิ ◌ี ◌ึ ◌ื ◌ุ ◌ู<br>เ แ โ ใ ไ ๅ ฤ ฦ |
| Tone marks | 4 | ◌่ ◌้ ◌๊ ◌๋ |
| Diacritics | 5 | ◌ฺ ◌์ ◌็ ◌ํ ◌๎ |
| Numerals | 10 | ๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙ |
| Other symbols | 6 | ๆ ฿ ฯ ๏ ๚ ๛ |
| Total | 87 | |

Figure 4. All 87 Thai scripts consist of 44 consonants, 18 vowels, 4 tone marks, 5 diacritics, 10 Thai numerals, and 6 special symbols

### A. Dataset

Thai has been an official language longer than 720 years [1] in Siam since King Ramkhamhaeng of Sukothai Kingdom. Thai has 87 scripts [3], including 44 consonants, 18 vowels, 4 tone marks, 5 diacritics, 10 Thai numerals, and 6 special symbols (as shown in Fig. 4). These scripts can be written in different vertical positions; in both Thai handwriting [56] and printed text recognition [57].

For the data preparation, physical Thai handwritten real images are first resized and/or divided into 512x128 pixels in width and height. The collection contains 8,324 prepared Thai handwritten images with 87 target classes as the primary data. The ratio of training and validation is partitioned as 90:10.

### B. Network and Parameter Settings

In generator (G), the pile of noise is split into a vector that its size depends on the length of Thai textual content. The dimension of fully-connected layers is factorized into 256x2x4. All Thai handwritten images with their noises are input to each up-sampling residual block (ResBlock) based on bi-linear interpolation [55]. ResBlocks have the different number of filters: 256, 128, 128, 64, 64, 32, 16, 16 (as a collection of kernels) through the 25 units of Conditional batch normalization (CBN) with the activation function by Rectified linear unit (ReLU) [58]. Inversely, the down-sampling residual blocks (ResBlocks, based on average pooling [52]) also have different filters: 16, 16, 32, 64, 128, 128, 256 through general Batch normalization (BN) with the LeakyReLU as the activation function for ResBlocks in discriminator (D).

### C. Metrics

*1) Realistic measurement:* To evaluate the realistic of proposed algorithm based on generative model (both generator and discriminator), Fréchet Inception Distance (FID) [59] is a metric for measuring similarity between real and generated images as two different multivariate Gaussians as (6). For the evaluation, the lower value means higher performance.

$$FID = \left\|\mu(X) - \mu(X_{generated})\right\|_2^2 + Tr\left(\sum_X + \sum_{X_{generated}} - 2\times\sqrt{\sum_X \sum_{X_{generated}}}\right) \tag{6}$$

where $\mu(X)$ and $\mu(X_{generated})$ as average of real images and generated images, $\|\bullet\|_2^2$ as Euclidean $L_2$ normalization, $\sum_X$ and $\sum_{X_{generated}}$ as covariance matrices of real images and generated images, $Tr(\bullet)$ as the main diagonal of a matrix

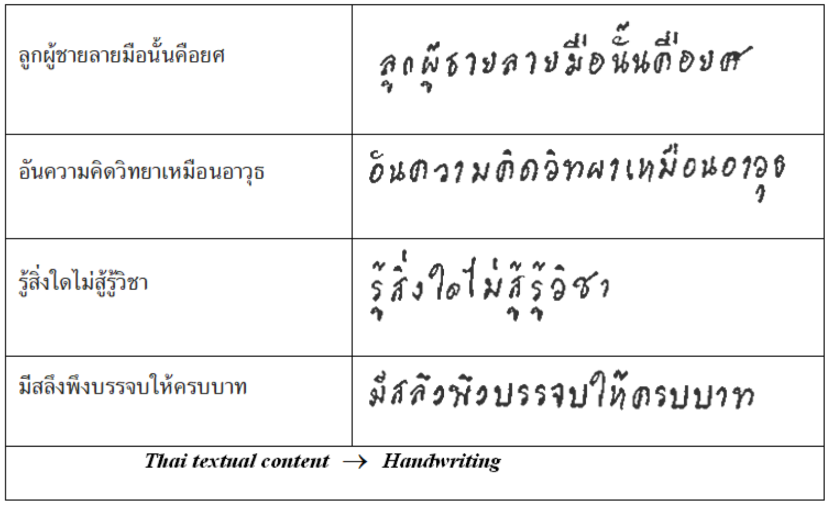


Figure 5. Given any Thai textual content/information, Thai handwriting generation by ThaiWritableGAN. (Note that these quotes are from Thai literatures written by Sunthorn Phu – honored by UNESCO [4-5] as one of the world's great poets.)

*1) Character-level recognition:* To evaluate the character-level Thai handwritten recognition, Word Error Rate (WER) [60] – a well-known metric in machine translation is used to measure a number of miscalculated scripts that are not written by ThaiWritableGAN ( $I$ ), a number of scripts written by ThaiWritableGAN that were lost ( $D$ ) and a number of scripts were substituted in the original content ( $S$ ), respectively.

$$WER = \left( \frac{I + D + S}{N} \right) \times 100 \quad (7)$$

where $I$ , $D$ and $S$ are divided by a number of original scripts ( $N$ ). WER was used as metric for Thai script recognition [61] and also in this composing Thai handwritten content.

### D. Results and Discussion

Some generated Thai handwriting examples (based on well-known Thai quotes) by ThaiWritableGAN are shown in Fig. 5. For scientific knowledge discovery, the major findings can be categorized into 5 main experimental issues.

*1) Realistic improvement by Self-attention mechanism:* Since self-attention as a core sequence-to-sequence mechanism [54] in Transformer that can be used to Thai character-level embedding, self-attention works well on not only coreference resolution and vanishing gradient but also weight well-defined in Thai handwritten data. Self-attention is included in both generator and discriminator, instead of the flat CNN-RNN architecture. Hence, generative model with self-attention provides more realistic samples for almost 3 times than without self-attention, as TABLE I.

TABLE I. SELF-ATTENTION IMPROVEMENT

| **Self-attention** | **FID** |
|---|---|
| Included | 27.86 |
| Not included | 79.98 |

*2) Gradient balancing:* The parameter $\alpha$ in (4) refers to how recognizer gradient ( $\|\nabla R\|$ ) higher than discriminator gradient ( $\|\nabla D\|$ ). The different parameter $\alpha$ values impact the realistic Thai-handwritten samples as shown in TABLE II. Given $\alpha$ =0.1, $\|\nabla R\|$ is smaller than $\|\nabla D\|$ which means CTC cost set to be much less significance than discrimination. In contrast, the $\alpha$ is set to be 10 (as $\|\nabla R\| > \|\nabla D\|$ ) that generates lower realistic quality than the $\alpha$ =0.1 setting. The best fit of $\alpha$ finally equals as 1 that it can generate the most realistic samples.

TABLE II. GRADIENT BALANCING BETWEEN $\|\nabla R\|$ AND $\|\nabla D\|$

| $\alpha$ | **FID** |
|---|---|
| 0.1 | 86.37 |
| 1 | 27.86 |
| 10 | 257.43 |

TABLE III. COMPARISON BETWEEN (4) AND (5)

| Optimization | FID | WER |
|---|---|---|
| $\nabla R \leftarrow \alpha\left(\frac{\sigma(\nabla D)}{\sigma(\nabla R)}(\nabla R-\mu(\nabla R))+\mu(\nabla D)\right)$ (as (4)) | 27.86 | $17.48 \pm 0.38$ |
| $\nabla R \leftarrow \alpha\left(\frac{\sigma(\nabla D)}{\sigma(\nabla R)}\bullet\nabla R\right)$ (as (5)) | 29.51 | $14.72 \pm 0.47$ |

*3) Comparison between Eq.(4) and (5):* The full version in (4) considers the average in discriminator and recognizer gradients ( $\mu(\nabla D)$ and $\mu(\nabla R)$ ) as well as standard deviation $\sigma(\nabla D)$ and $\sigma(\nabla R)$ that provide higher FID. Although $\mu(\nabla D)$ and $\mu(\nabla R)$ boost the performance in realistic generation, they seem to be noise for Thai handwritten recognition (as shown in TABLE III).

*4) Realistic improvement by recognizer:* The recognizer (R) boosts the readability of synthesized Thai handwriting that is generated by Generator (G). The additional R is added in the proposed Generative adversarial network (GAN) architecture that can be used as the recognition penalty for the readable handwriting generation (as the optimization for G). To combine R with discriminator (D), the handwritten samples become better realistic performance, as TABLE IV.

TABLE IV. REALISTIC IMPROVEMENT BY RECOGNIZER

| $L_D$ | $L_R$ | FID |
|---|---|---|
| ✓ | × | 58.91 |
| ✓ | ✓ | 27.86 |

5) *Comparison between adversarial losses:* Thai handwriting generation is tested in different architectures: GAN [11], Wasserstein GAN [62], Least Square GAN [63], Loss Sensitive GAN [64] and Geometric GAN [12], respectively.

TABLE V. COMPARISON BETWEEN ADVERSARIAL LOSSES

| Adversarial Loss | FID |
|---|---|
| GAN | 49.84 |
| Wasserstein GAN | 92.08 |
| Least Square GAN | 80.61 |
| Loss Sensitive GAN | 33.27 |
| Geometric GAN | 27.86 |

From Table V, the most stable training for Thai handwriting generation is Geometric GAN that uses Support vector machine (SVM)'s hyper-plane to separate between real and generated samples [12]. The realistic image can be estimated by the maximum margins between them (as well as the concept of Loss Sensitive GAN). However, Wasserstein GAN and Least Square GAN lead to the style collapse problem. The well-defined loss function and/or architecture can totally boost generator (G) to generate more realistic handwriting images.

## 5. CONCLUSION

As it relates to the joint edge between Generative adversarial network (GAN) and Thai handwriting recognition as a new research topic in Thai language with computer vision, this paper firstly proposes a generative model (named "ThaiWritableGAN") to cope with this problem. For the novelty, the proposed unsupervised model consists of a generator (G) – to generate Thai handwritten samples; discriminator (D) – to classify that the Thai handwritten image is real or generated; recognizer (R) – to imply realistic and readable image, respectively. The physical handwritten images are collected as primary data. Fréchet Inception Distance (FID) is used to evaluate the similarity between real and generated data. The character-level Thai handwritten recognition is measured by Word Error Rate (WER). For the major findings, self-attention and additional recognizer (R) boosts more realistic image generation. The fittest gradient balancing between discriminator and recognizer is 1. For two different optimization settings, the optimization Eq.(4) provides more realistic output but less recognition rate than Eq.(5). For future work, the model may compose the content in many writing styles. To apply other new GAN extensions that may provide more written styles beyond the realistic and readable quality against vanishing gradient in the long range of Thai character-level. It is possible to have a Thai handwriting generation contest that allows Thai researchers to propose their GAN-based algorithms for the beauty of Thai calligraphy competition.

## ACKNOWLEDGMENT

For introducing a new challenge of Thai handwriting generation, the paper "ThaiWritableGAN: Handwriting Generation under Given Information" was extended from the unsupervised Generative adversarial network (GAN). Thanks to the expert from Thailand's National Electronics and Computer Technology Center (NECTEC) for the technical guidance. In order to Thai heritage conservation, this paper was proposed to apply computer vision to Thai as our Rajabhat's mission. Dedicated to Chandrakasem Rajabhat University, all physical Thai handwritten images in the experiment were watermarked as the ownership of local primary data (referred to this paper) that could be requested by the authors' emails.

## REFERENCES


[1] L. Inthajakra, A. Prachyapruit and S. Chantavanich, "The Emergence of Communication Intellectual History in Sukhothai and Ayutthaya Kingdom of Thailand," *Social Science Asia,* vol. 2, no. 4, pp. 32-41, 2016.

[2] P. Mookdarsanit and L. Mookdarsanit, "TGF-GRU: A Cyber-bullying Autonomous Detector of Lexical Thai across Social Media," *NKRAFA Journal of Science and Technology,* vol. 15, no. 1, pp. 50-58, 2019.

[3] H. T. Koanantakool, T. Karoonboonyanan and C. Wutiwiwatchai, "Computers and the Thai Language," *IEEE Annals of the History of Computing,* vol. 31, no. 2, pp. 46-61, 2009.

[4] K. Pimpuang, "How Sunthorn Phu Expressed his Aesthetic Values through Sanskrit*?," Journal of Language and Linguistics,* vol. 34, no. 2, pp. 46-62, 2016.

[5] K. Champeon, "Thailand's Shakespeare? Sunthorn Phu," 2003. [Online]. Available: http://thingsasian.com/story/thailands-shakespeare-sunthorn-phu. Accessed: October 8, 2020.

[6] P. Suklek and P. Anunsiriwat, "Linguistic Strategies of Otherness Creation in Nirat Poetry of Soonthonphu," *Humanities and Social Sciences Review,* vol. 6, no. 2, pp. 16-34, 2018.

[7] D. Cooper, "How Do Thais Tell Letters Apart?," in *the 1996 International Symposium on Language and Linguistics,* Mahidol University, Thailand, 1996.

[8] C. Tanprasert and T. Koanantakool, "Thai OCR: A Neural Network Application," in *the 1996 Digital Processing Applications,* pp.90-95, Australia, 1996.

[9] V. Sornlertlamvanich T. Potipiti, C. Wutiwiwatchai and P. Mittrapiyanuruk, "The State of the Art in Thai Language Processing," in *the 2000 Annual Meeting on Association for Computational Linguistics,* pp.1-2, 2000.

[10] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based Learning applied to Document Recognition," *Proceeding of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Nets," in *the 2014 International Conference on Neural Information Processing Systems (NIPS '14),* pp.2672-2680, 2014.

[12] J. H. Lim and J. C. Ye, "Geometric GAN," 2017. [Online]. Available: https://arxiv.org/pdf/1705.02894.pdf. Accessed: October 8, 2020.

[13] C. Kimpan and S. Walairacht, "Thai Characters Recognition," in *the Symposium on Natural Language Processing (SNLP '94),* Thailand, 1994.

[14] C. Lursinsap and C. Khunasaraphan, "Simulated Light Sensitive Model for Handwritten Digit Recognition," in *IJCNN International Joint Conference on Neural Networks (IJCNN '92),* pp.13-18, 1992.

[15] C. Tanprasert, W. Sinthupinyo and P. Dubey, "On the Printed Thai Optical Character Recognition," in *the 1996 National Electronics and Computer Technology (NECTEC),* Thailand, 1996. [in Thai].

[16] S. Meknavin and B. Kijsirikul, "Combining Trigram and Winnow in Thai OCR Error Correction," *NECTEC Technical Journal,* vol. 1, no. 1, pp. 19-26, 1999.

[17] C. Tanprasert, W. Sinthupinyo, P. Dubey, T. Tanprasert, "Improved Mixed Thai & English OCR using Two-step Neural Net Classification," *NECTEC Technical Journal,* vol. 1, no. 1, pp. 41-46, 1999.

[18] A. Kawtrakul and P. Waewsawangwong, "Multi-feature Extraction for Printed Thai Character Recognition," in *the 2000 Symposium on Natural Language Processing (SNLP '00)*, 2000.

[19] W. Kasemsiri and C. Kimpan, "Printed Thai Character Recognition using Fuzzy-Rough Sets" in *the 2001 IEEE Region 10 International Conference on Electrical and Electronic Technology,* pp.326-330, 2001.

[20] P. Phokharatkul and P. Pantaragphong, "Printed Thai Characters Recognition Using Rough Sets and Fractal Dimension," in *International Symposium on Nonlinear Theory and Its Applications*, 2002.

[21] B. Kijsiriul, , S. Sinthupinyo and A. Supanwansa, "Thai Printed Character Recognition by Combining Inductive Logic Programming with Backpropagation Neural Network," in *Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems Microelectronics and Integrating Systems,* pp.539-542, 1998.

[22] U. Marang, P. Phokharatkul and C. Kimpan, "Recognition of Printed Thai Characters using Boundary Normalization and Fuzzy Neural Networks," in *the 2000 Symposium on Natural Language Processing (SNLP '00)*, 2000.

[23] U. Suttapakti, K. Woraratpanya and K. Pasupa, "Font Descriptor Construction for Printed Thai Character Recognition," in *the 2013 IAPR International Conference on Machine Vision Applications*, 2013.

[24] T. Siriteerakul, V. Boonjing and R. Gullayanon, "Character Classification Framework based on Support Vector Machine and k-nearest Neighbour Schemes," *Science Asia,* vol. 42, no. 1, pp. 46-51, 2016.

[25] I. Methasate and S. Marukatut, "BEST 2013: Thai Printed Character Recognition Competition," in *the 2013 Symposium on Natural Language Processing (SNLP '13)*, 2013.

[26] K. Kesorn and P. Phawapoothayanchai, "Optical Character Recognition (OCR) enhancement using an approximate string matching technique," *Engineering and Applied Science Research,* vol. 45, no. 4, 2018.

[27] S. Chaiwatanaphan, C. Pluempitiwiriyawej and S. Wangsiripitak, "Printed Thai Character Recognition using Shape Classification in Video Sequence along a Line," *Engineering Journal,* vol. 21, no. 6, pp. 37-45, 2017.

[28] S. Marukatut, "BEST 2019," 2019. [Online]. Available: https://medium.com/nectec/best-2019-58e6d2197ad8. Accessed: October 8, 2020. [in Thai].

[29] C. Pornpanomchai, D. N. Batanov and N. Dimmitt, "Recognizing Thai Handwritten Characters and Words for Human–computer Interaction," *International Journal of Human-Computer Studies,* vol. 55, no. 3, pp. 259-279, 2001.

[30] A. Pornchaikajornsak and A. Thammano, "Handwritten Thai character recognition using fuzzy membership function and fuzzy ARTMAP," in *the2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium,* pp.40-44, 2003.

[31] P. Phokharatkul and C. Kimpan, "Handwritten Thai Character Recognition Using Fourier Descriptors and Genetic Neural Networks," *Computational Intelligence*, vol. 18, no. 3, pp. 270-293, 2002.

[32] C. Pornpanomchai, V. Wongsawangtham, S. Jeungudomporn and N. Chatsumpun, "Thai Handwritten Character Recognition by Genetic Algorithm (THCRGA)," *IACSIT International Journal of Engineering and Technology*, vol. 3, no. 2, pp. 148-153, 2011.

[33] R. Nopsuwanchai, A. Biem and W. F. Clocksin, "Maximization of Mutual Information for Offline Thai Handwriting Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1347-1351, 2006.

[34] I. Methasate and S. Sae-tang, "The Clustering Technique for Thai Handwritten Recognition," in *the 2004 Int'l Workshop on Frontiers in Handwriting Recognition*, pp.450-454, 2004.

[35] S. Chunwijitra, P. Woraratpanya, S. Nitsuwat and P. Nilsuk, "Thai Online Handwriting System Using Neural Networks and Fuzzy Logic," *Information Technology Journal*, vol. 1, no. 2, pp. 77-82, 2005.

[36] C. Karnchanapusakij, P Suwannakat, W. Rakprasertsuk and N. Dejdumrong, "Online Handwriting Thai Character Recognition," in *the 2009 International Conference on Computer Graphics, Imaging and Visualization*, pp.323-328, 2009.

[37] P. Jamjuntr and N. Dejdumrong, "An Improved Online Handwriting Thai Character Recognition," in *the 2011 International Conference on Computer Graphics, Imaging and Visualization*, pp.87-91, 2011.

[38] E. Chaowicharat, N. Cercone and K. Naruedomkul, "Novel Curve Signatures and a Combination Method for Thai On-Line Handwriting Character Recognition," in *the 2012 International Florida Artificial Intelligence Research Society Conference*, 2012.

[39] F. J. J. Joseph and P. Anantaprayoon, "Offline Handwritten Thai Character Recognition Using Single Tier Classifier and Local Features," in *the 2018 International Conference on Information Technology*, pp.1-4, 2018.

[40] O. Surinta, M. F. Karaaba, L. R. B. Schomaker and M. A. Wiering, "Recognition of handwritten characters using local gradientfeature descriptors," *Engineering Applications of Artificial Intelligence*, vol. 15, no. 1, pp. 405-414, 2015.

[41] P. Mookdarsanit and L. Mookdarsanit "ThaiWrittenNet: Thai Handwritten Script Recognition using Deep Neural Networks," *Azerbaijan Journal of High Performance Computing*, vol. 3, no. 1, pp.75-93, 2020.

[42] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *the 2012 International Conference on Neural Information Processing Systems (NIPS '12),* 2012.

[43] S. Gonwirat and O. Surinta, "Improving Recognition of Thai Handwritten Characters with Deep Convolutional Neural Networks," in *the 2020 International Conference on Information Science and System*, pp.82-87, 2020.

[44] R. Chamchong, W. Gao and M. D. McDonnell, "Thai Handwritten Recognition on Text Block-Based from Thai Archive Manuscripts," in *the 2019 International Conference on Document Analysis and Recognition (ICDAR '19),* pp.1346-1351, 2019.

[45] T. Lapjaturapit, K. Viriyayudhakorn and T. Theeramunkong, "Multi-Candidate Word Segmentation using Bi-directional LSTM Neural Networks," in *the 2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems*, pp.1-6, 2018.

[46] A. Graves, "Generating Sequences With Recurrent Neural Networks," 2013. [Online]. Available: https://arxiv.org/pdf/1308.0850.pdf. Accessed: October 8, 2020.

[47] P P. Isola, J-Y. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '17),* pp.5967-5976, 2017.

[48] Y. Tian, "zi2zi: Master Chinese Calligraphy with Conditional Adversarial Networks" 2019. [Online]. Available: https://github.com/kaonashi-tyc/zi2zi. Accessed: October 8, 2020. [only the zi2zi's code available on github].

[49] B. Chang, Q. Zhang, S. Pan and L. Meng, "Generating Handwritten Chinese Characters using CycleGAN," in *the 2018 IEEE Winter Conference on Applications of Computer Vision*, pp.199-207, 2018.

[50] S. Fogel, H. Averbuch-Elor, S. Cohen, S. Mazor and R. Litman, "ScrabbleGAN: Semi-Supervised Varying Length Handwritten Text Generation," in *the 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '20),* pp.4324-4333, 2020.

[51] L. Kang, P. Riba, Y. Wang, M. Rusinol, A. Fornes and M. Villegas, "GANwriting: Content-Conditioned Generation of Styled Handwritten Word Images," 2020. [Online]. Available: https://arxiv.org/pdf/2003.02567.pdf. Accessed: October 8, 2020.

[52] E. Alonso, B. Moysset and R. Messina, "Adversarial Generation of Handwritten Text Images Conditioned on Sequences," in *the 2019 International Conference on Document Analysis and Recognition (ICDAR '19),* pp.481-486, 2019.

[53] M. Raghu and E. Schmidt, "A Survey of Deep Learning for Scientific Discovery," 2020. [Online]. Available: https://arxiv.org/pdf/2003.11755.pdf. Accessed: October 8, 2020.

[54] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention Is All You Need," in *the 2017 International Conference on Neural Information Processing Systems (NIPS '17),* 2017.

[55] Z. Wang, J. Chen and S. C. H. Hoi, "Deep Learning for Image Super-resolution: A Survey," 2019. [Online]. Available: https://arxiv.org/pdf/1902.06068.pdf. Accessed: October 8, 2020.

[56] T. Phienthrakul and M. Samnienggam, "Comparative Results of Attribute Reduction Techniques for Thai Handwritten Recognition with Support Vector Machines," in *the 2016 International Conference on Computing and Information Technology (ICCIT '16),* pp.67-77, 2016.

[57] T. Emsawas and B. Kijsirikul, "Thai Printed Character Recognition using Long Short-Term Memory and Vertical Component Shifting," in *the 2016 Pacific Rim International Conference on Artificial Intelligence (PRICAI '16)*, pp. 106-115, 2016.

[58] H. H. Kim, "Test Case Generation for Convolutional Neural Network," *International Journal of Computing and Digital Systems,* vol. 9, no.2, pp. 271-280, 2020.

[59] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler and S. Hochreiter, "Gans Trained by A Two Time-scale Update Rule Converge to A Local Nash Equilibrium," in *the 2017 International Conference on Neural Information Processing Systems (NIPS '17)*, pp. 6629–6640, 2017.

[60] A. Marzal and E. Vidal, "Computation of Normalized Edit Distance and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 926-932, 1993.

[61] M. Popović and H. Ney, "Word Error Rates: Decomposition over POS Classes and Applications for Error Analysis," in *the 2007 Workshop on Statistical Machine Translation*, pp. 48-55, 2007.

[62] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017. [Online]. Available: https://arxiv.org/pdf/1701.07875.pdf. Accessed: October 8, 2020.

[63] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, "Least Squares Generative Adversarial Networks," in *the 2017 IEEE International Conference on Computer Vision (ICCV '17)*, pp. 2813– 2821.

[64] G-J. Qi, "Loss-sensitive Generative Adversarial Networks on Lipschitz Densities," 2017. [Online]. Available: https://arxiv.org/pdf/1701.06264.pdf. Accessed: October 8, 2020.


**LAWANKORN MOOKDARSANIT** is an assistant professor in information technology at business information systems (BISS), faculty of management science, Chandrakasem Rajabhat University, Bangkok, Thailand. She received her bachelor degree (business computer) in 2009, master degree (information technology) in 2011, and Ph.D. degree (information technology) in 2016, respectively. Her current administrative position is an assistant to the dean for short-course curriculum design, faculty of management science, Chandrakasem Rajabhat University. Her research interests include object detection, image classification and image synthesis. According to Rajabhat's mission, she is also interested in machine learning and data mining with social applications as the solution for local problems, e.g., automatic agricultural quality verification by image, HR intelligence and hate speech detection.

**PAKPOOM MOOKDARSANIT** received his Ph.D. information technology in 2017, M.S. computer science in 2013 and B.Eng. computer engineering in 2010. He is an assistant professor in information technology at computer science and artificial intelligence (CSAI), faculty of science, Chandrakasem Rajabhat University, Bangkok, Thailand. He is also an assistant director in the list of administrative board, Chandrakasem Rajabhat University – Chainat. His main research areas are image forensics, image recognition and natural language processing. His researching and coding skills are also to apply deep learning into various industrial applications: autonomous image captioning, food calorie estimation, bully detection, spamming text filtering, Turing test and reCaptcha images, etc.