The Cosmetic Surgery Recommendation: Facial Acne Localization and Recognition

Pakpoom Mookdarsanit[1], Lawankorn Mookdarsanit[2]

[1]Computer Science and Artificial Intelligence, Faculty of Science
Chandrakasem Rajabhat University
Bangkok, Thailand
pakpoom.m@chandra.ac.th

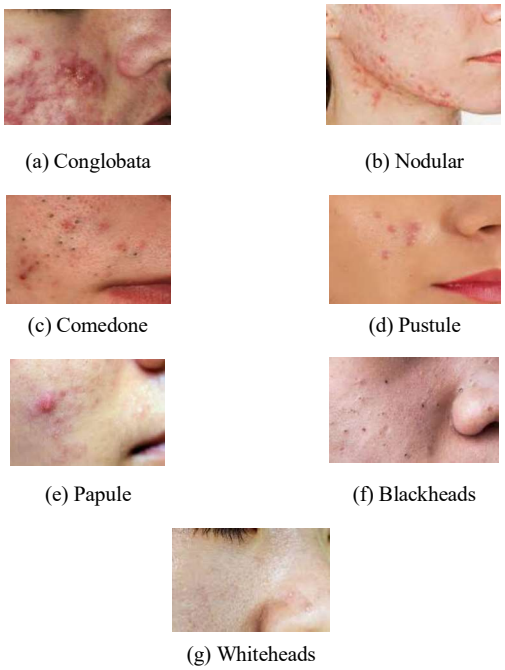
[2]Business Information Systems, Faculty of Management Science
Chandrakasem Rajabhat University
Bangkok, Thailand
lawankorn.s@chandra.ac.th

Abstract— Oftentimes, patients took their facial images to be recommended by the doctor about the cosmetic surgery across the official social media. The doctors could classified the skin disease type and remedy as the new normal. This paper introduced the acne localization and recognition from the patient's image as a facial surgery recommendation. The proposed system was based on "You only look once version 3 (YOLOv3)" that was fast and high correctness in both position detection and acne name recognition. This facial acne localization and recognition covered 7 different acne's types for the cosmetic surgety recommendation: (1) Conglobata, (2) Nodular, (3) Comedone, (4) Pustule, (5) Papule, (6) Blackheads and (7) Whiteheads. For the contribution, this showed a concrete practice to adopt deep learning in surgery industry. It is possible to use artificial general intelligence (AGI) to leverage the remedy recommendation process in cosmetic surgery industry by any AGI start-ups. ***(Abstract)***

Keywords-Skin Disease Detection; Facial Analytics; Cosmetic Surgery Recommendation; YOLO; Acne Detection (key words)

I. INTRODUCTION

Cosmetic surgery is one of health and medical industry that has been grown rapidly in Asia, especially in ladies. The purpose of surgery could be facelift, whitening, vanity, and wrinkle or acne removal. As to the new normal, the patients often sent their facial images to the clinic/hospital's official social media to ask for cosmetic surgery recommendation. To dive into those facial images, there were so many facial acne images to be considered by the doctor. It was possible to leverage this manual recommendation by the help of computer vision [1]. There were so many computer vision applications, e.g., food classification [2-5], gesture recognition [6-8], agricultural applications [9-11], wild scene text detection [12-13], tourism recommendation [14-16], culture and arts [17-20], remote sensing [21-22], and HR Intelligence [23]. Face analytic [24] was a type of computer vision that there were 4 normal applications: (1) face recognition, (2) facial landmark recognition, (3) face verification, and (4) face synthesis. Face recognition was proposed to find what facial image type was. To localize the eyes, ears, nose, and mouth were in facial landmark recognition. Face verification was differ from face recognition. The verification was to find whose facial image was. And face synthesis was to generate the new facial images by generative models. For the acne analytics, the literature could be categorized in skin detection [25], skin care product recommendation [26], or acne detection [27-28].



(a) Conglobata (b) Nodular (c) Comedone (d) Pustule (e) Papule (f) Blackheads (g) Whiteheads

Figure 1. The 7 facial acne types in this paper

For the contribution, this paper extended the previous acne analytics coupled with face recognition to design the acne localization and recognition through the facial image

to automatically detect the position of acnes; and classify what name of each acne was. This system covered 7 acne's types, e.g., (1) Conglobata, (2) Nodular, (3) Comedone, (4) Pustule, (5) Papule, (6) Blackheads and (7) Whiteheads, as shown in Fig.1. The proposed facial acne localization and recognition was based on "You only look once version 3 (YOLOv3)" [29] that was one-stage object detection. One-stage object detection did not compute the region, while two-stage object detection [30] (e.g., Faster R-CNN [31]) have the region of object. The correctness of one-stage object detection could be comparable to two-stage ones but faster.

Furthermore, this facial acne detection as one of computer vision could be coupled with natural language processing [32-35] to be applied text captioning [36] to generate text of image [37] and/or text to image methods (e.g., Imagen [38], Stable Diffusion [39], DALL-E [40-41], Make-a-scene [42]) to draw facial acne images from text by artificial general intelligence (AGI). For example, the AGI might get the facial acne image; and generate the facial image without acne to make patients see their faces after the remedy.

This paper was organized into 4 parts. The part 2 talked about facial acne localization and recognition. The acne detection results were in part 3. And part 4 was conclusion.

II. Facial Acne Localization and Recognition

The proposed facial acne detection (localization and recognition) is based on "You only took once version 3 (YOLOv3)" [29] that is such a one-stage object detection.YOLOv3 is extended batch normalization (BatchNorm) [43] for feature modification and k-means algorithm for bounding box with dimension prior from YOLOv2 and also added Residual Network (ResNet) [44] for skip connection, as shown in Fig.2.

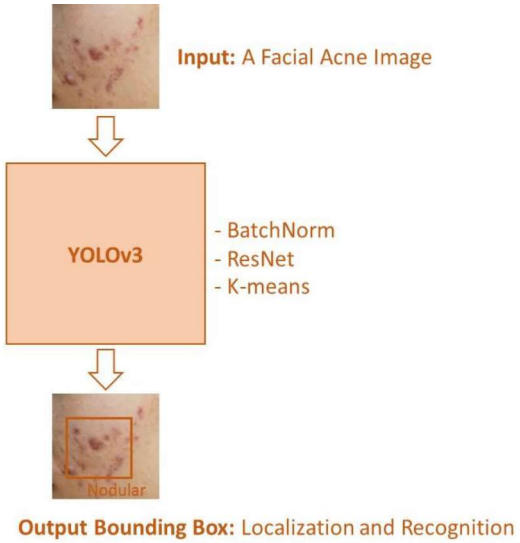


Figure 2. Facial acne localization and recognition based on YOLOv3

A. Acne Detection

The facial acne detection by YOLO consists of localization and recognition that the loss function (L_{Acne}) can be defined by (1)

$$L_{Acne} = L_{Localization}\left(B_a, B_p\right) + L_{\text{Re}cognition}\left(y, \hat{y}\right) \quad (1)$$

where B_a, B_p are the actual bounding box and predicted one, $L_{Localization}(\bullet)$ is the localization loss, $y, \hat{y}$ are the actual acne type and predicted acne type, and $L_{\text{Re}cognition}(\bullet)$ is the recognition loss.

The composition of basic YOLO [45] is bounding box, intersection over union, anchor box and non-max suppression (NMS).

1) Bounding Box refers to the localization box as the output of objects' positions where are b_w, b_h width and length of the box, and b_x, b_y the first pixel position of the target acne object within the image in (x, y)

2) Intersection over Union ($IoU\left(B_a, B_p\right)$) is a ratio function to measure the the intersection of B_a, B_p per the union of B_a, B_p as in (2)

$$IoU\left(B_a, B_p\right) = \frac{B_a \cap B_p}{B_a \cup B_p} \quad (2)$$

3) Anchor Box is the box selection of the interesting objects.

4) Non-max Suppression (NMS) is technique to combine many common areas from different anchor boxes into one, as in Fig.3.

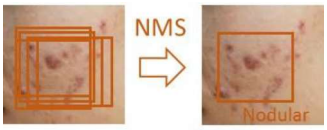


Figure 3. Non-max suppression (NMS)

B. Batch Normalization

The basic normalization [43] is proposed to make the center of data is at $(0,0)$, called "zero center", in Fig.4. Batch Normalization (BatchNorm) is to adapt the input feature value into the zero center before processing by neural network.

$$\hat{x}_{i,j} = \frac{x_{i,j} - \mu_j}{\sqrt{\sigma^2 + \varepsilon}} \quad (3)$$

where $\hat{x}_{i,j}, x_{i,j}$ are new feature value and old value, μ_j, σ^2 the average and variance of all features in the

same batch, i, j the order of dataset and batch, and ε is the bias

The dimensional output of BatchNorm ($d_{i,j}$) can be defined by (4), where $\gamma = \sigma$ and $\beta = \mu$

$$d_{i,j} = \frac{\gamma(x_{i,j} - \mu_j)}{\sigma^2 + \beta} \quad (4)$$

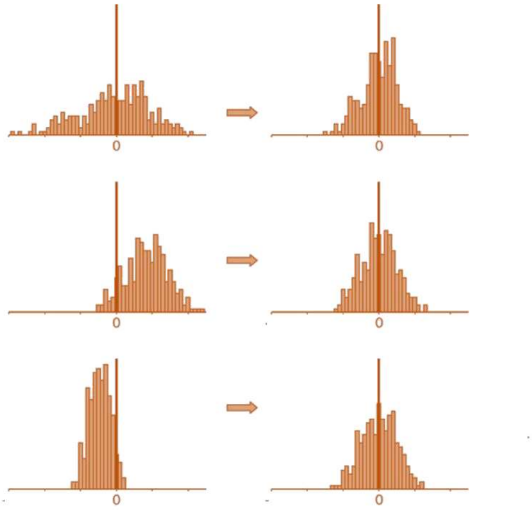


Figure 4. The concept of batch normalization (BatchNorm)

C. Residual Network

Residual Network (ResNet) [44] was the winner of ImageNet Large Scale Visual Recognition Challenge 2015 (ILSVRC 2015). The key concept of ResNet is skip connection, coupled with element-wise addition and BatchNorm with ReLU activation function.

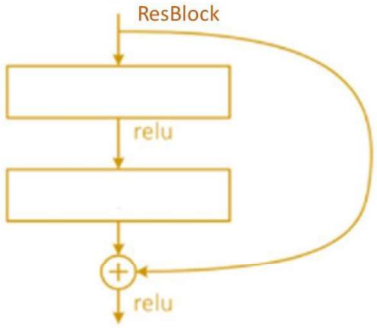


Figure 5. The architecture of residual network (ResNet)

D. K-means Algorithm

K-means algorithm is an un-supervised clustering technique to group all features into k groups by considering distance or similarity. The 2 different features having the less distance should be grouped into the same group (the more distance grouped in the different group). In contrast, the 2 ones having the more similarity should be grouped into the same group (the less similarity grouped in the different group). YOLOv2 (or YOLO9000) [45-46] and YOLOv3 [29] apply K-means for bounding box with dimension prior.

III. The Acne Detection Results

The results of facial acne detection could be divided into localization and recognition. The annotation software used was LabelIMG [47] for adding the acne type to each image. And the YOLOv3 learning those images with annotations ran on Google Colab (based on the version 3 available in TensorFlow).

A. Experimental Settings

All 2,304 facial acne images were firstly combined and that covered 7 acne types: (1) Conglobata, (2) Nodular, (3) Comedone, (4) Pustule, (5) Papule, (6) Blackheads and (7) Whiteheads. Since the privacy of person identification and computational reduction, some facial landmarks (e.g., hair, ears, eyes, nose, mouth) were reduced.

Since the proposed system was a full-supervised learning that needed both data with annotation (or labeling), all 1,920 acne images were manually annotated the acne types using LabelIMG [47]; and stored in training folder. And all acne types were stored in the format of xml file. The 234 images were stored in validation one. Each folder had annotations and images.

To model the facial acne detection, the YOLOv3 learnt all acne types from the image data with annotations that ran on Colab cloud; the loss acceptance in (1) was set as equal or higher than 60 ($L_{Acne} \geq 0.6$). The ratio between training and validation was 80:20.

B. Acne Detection Results

Some experimental results were shown in Fig.6. The overall detection (localization and recognition) was in the high level of correctness. The proposed facial localization and recognition can be a cosmetic surgery recommendation on the 7 acne type domain in order to the medical online service.

TABLE I. The Facial Acne Localization Comparrsion

One-stage Object Detection	**Localization** (mAP)	**Speed** (FPS)
YOLO9000 [46]	31.6	42
SSD [48]	39.4	63
YOLOv3 [29]	42.2	58

For the acne localization, the mean average precision (mAP) and flop per second (FPS) were used and compared to other methods. The localization result found that

YOLOv3 [29] was the highest mAP (under the $IoU \geq 0.5$) and also the acceptable speed in term of FPS. Since the mAP of Single-shot Detector (SSD) [48] could be compared to YOLOv3, it seemed take more processing time in term of FPS. Meanwhile, YOLO9000 [46] was faster than YOLOv3 but the localization correctness was very low and unacceptable, as shown in Table I.

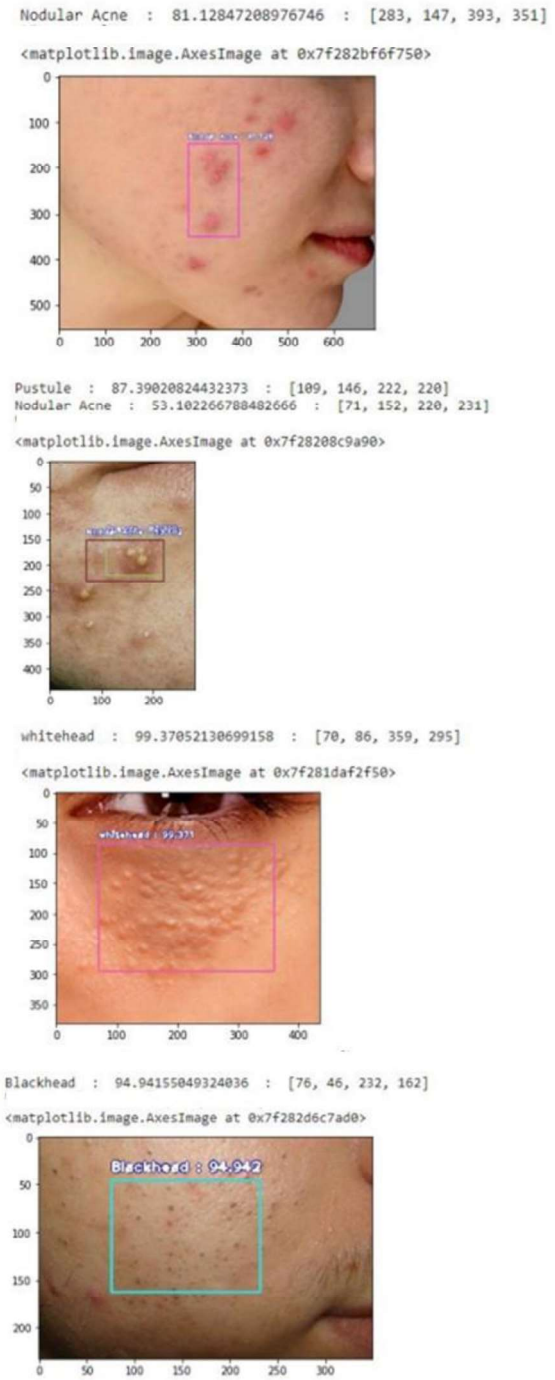


Figure 6. Some experimental results

For the recognition, the overall recognition in 7 facial acne types was shown in Fig.7. The Whiteheads was the highest accuracy since it was physical unique. While Comedone was only 0.74 of accuracy, this might be the resolution and appearance, and less number of samples.

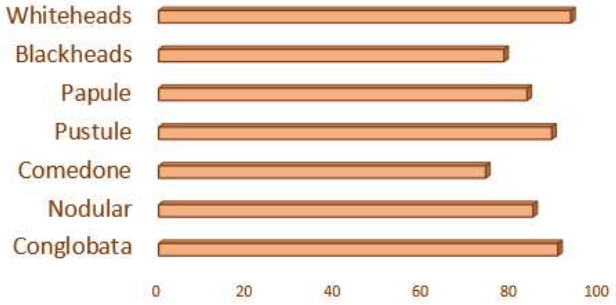


Figure 7. The 7 acne types recognition results in accuracy

IV. Conclusion

As referred to the automatic recommendation on cosmetic surgery, this paper introduced the facial acne localization and recognition based on "You took only once version 3 (YOLOv3)". There were 2,304 facial acne images that covered 7 different acne's types, e.g., (1) Conglobata, (2) Nodular, (3) Comedone, (4) Pustule, (5) Papule, (6) Blackheads and (7) Whiteheads, respectively. Acne localization results were measured by mean average precision (mAP) and flop per second (FPS), while the recognition was done by accuracy.

For future extensions, this facial acne detection could be coupled with natural language processing (NLP) to be applied the automatic text captioning of image. Moreover, text to image (or image synthesis) methods (e.g., Imagen, Stable Diffusion, DALL-E, Make-a-scene) can be adopted to synthesize facial acne-reduced images from a raw text to make patients see their faces after the remedy easily.

Acknowledgment

For introducing a new intelligent acne detection from a facial image, this paper was adopted "You took only once version 3 (YOLOv3)" to localize the position of acne and recognize the acne's type. Thanks to the experts who help to combine the facial acne data. This work was proposed to make the local community and industry as a local data to be designed the research problem.

References

[1] H-T. Chan, T-Y. Lin, S-C. Deng, C-H. Hsia, and C-F. Lai, " Smart Facial Skincare Products Using Computer Vision Technologies," The 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Tokyo, Japan, 2021.

[2] L. Soimart and P. Mookdarsanit, "Ingredients estimation and recommendation of Thai-foods," in SNRU Journal of Science and Technology, vol.9, no.2, pp.509-520, 2017.

[3] P. Mookdarsanit and L. Mookdarsanit, "Name and Recipe Estimation of Thai-desserts beyond Image Tagging," in Kasembundit Engineering Journal, vol.8, Special Issue, pp.193-203, May. 2018.

[4] S. Turmchokkasam and K. Chamnongthai, "The Design and Implementation of an Ingredient-Based Food Calorie Estimation System Using Nutrition Knowledge and Fusion of Brightness and Heat Information," in IEEE Access, vol. 6, pp. 46863-46876, 2018.

[5] P. Mookdarsanit and L. Mookdarsanit, "The Autonomous Nutrient and Calorie Analytics from a Thai Food Image," in Journal of Faculty Home Economics Technology RMUTP, vol.8, no.1, pp.1-12, . 2020.

[6] P. Sharma and N. Sharma, "Gesture Recognition System," The 4th International Conference on Internet of Things: Smart Innovation and Usages, Ghaziabad, India, 2019, pp. 1-3.

[7] P. Mookdarsanit, et al., "A Content-based Image Retrieval of Muay-Thai Folklores by Salient Region Matching," in International Journal of Applied Computer Technology and Information Systems, vol.7, no.2, pp.21-26, 2018.

[8] P. Mookdarsanit and L. Mookdarsanit, "An Automatic Image Tagging of Thai Dance's Gestures," Joint Conference on ACTIS & NCOBA, Ayutthaya, Thailand, January 2018, pp. 76-80.

[9] C. Yang and H. Wei, "Plant Species Recognition Using Triangle-Distance Representation," in IEEE Access, vol. 7, pp. 178108-178120, 2019.

[10] L. Mookdarsanit and P. Mookdarsanit, "Thai Herb Identification with Medicinal Properties Using Convolutional Neural Network," in Suan Sunandha Science and Technology Journal, vol. 6, no.2, pp. 34-40, 2019.

[11] P. Mookdarsanit and L. Mookdarsanit, "PhosopNet: An Improved Grain Localization and Classification by Image Augmentation," in TELKOMNIKA Telecommunication, Computing, Electronics and Control, vol.19, no.2, pp. 479-490, 2021.

[12] T. Kobchaisawat and T. H. Chalidabhongse, "Thai Text Localization in Natural Scene Images using Convolutional Neural Network," The 2014 Signal and Information Processing Association Annual Summit and Conference, Siem Reap, Cambodia, 2014.

[13] L. Mookdarsanit and P. Mookdarsanit, "Combating the Hate Speech in Thai Textual Memes," in Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, no. 3, pp.1493-1502, 2021.

[14] P. Mookdarsanit and M. Rattanasiriwongwut, "Location Estimation of a Photo: A geo-signature MapReduce Workflow," in Engineering Journal, vol.21, no.3, pp.295-308, May 2017.

[15] P. Mookdarsanit and L. Mookdarsanit, "Contextual Image Classification towards Metadata Annotation of Thai-tourist Attractions," in ITMSoc Transactions on Information Technology Management, vol.3, no.1, pp. 32-40, 2018.

[16] L. Soimart and P. Mookdarsanit, "Name with GPS Auto-tagging of Thai-tourist Attractions from An Image," The 1017 Technology Innovation Management and Engineering Science International Conference, Nakhon Pathom, Thailand, 2017, pp. 211-217.

[17] P. Fugthong and P. Meesad, "Buddha Amulet Information Retrieval using Digital Images Combined with Feature Extraction and K-nearest Neighbor Techniques," in Information Technology Journal, vol.6, no. 2, pp. 34-40, 2013. [in Thai].

[18] L. Mookdarsanit, "The Intelligent Genuine Validation beyond Online Buddhist Amulet Market," in International Journal of Applied Computer Technology and Information Systems, vol.9, no.2, pp.7-11, 2019.

[19] M. Rattanasiriwongwut and P. Mookdarsanit, "MONTEAN Framework: A Magnificent Outstanding Native-Thai and Ecclesiastical Art Network," in International Journal of Applied Computer Technology and Information Systems, vol.6, no.2, pp.17-22, 2017.

[20] P. Mookdarsanit and M. Rattanasiriwongwut, "GPS Determination of Thai-temple Arts from a Single Photo," The 11th International Conference on Applied Computer Technology and Information Systems, Bangkok, January 2017, pp. 42-47.

[21] L. Soimart, et al., "The Segmentation of Satellite Image Using Transport Mean-shift Algorithm," 13th International Conference on IT Applications and Management, 2015, pp. 124-128.

[22] L. Soimart and M. Ketcham, "An efficient algorithm for earth surface interpretation from satellite imagery," in Engineering Journal, vol.20, no.5, pp.215-228, Nov. 2016.

[23] L. Mookdarsanit and P. Mookdarsanit, "The Insights in Computer Literacy toward HR Intelligence: Some Associative Patterns between IT Subjects and Job Positions," in Journal of Science and Technology RMUTSB, vol. 6, no. 2, pp. 12-23, 2020.

[24] L. Soimart, et al., "Gender Estimation of a Portrait: Asian Facial-significance Framework," The 6th International Conference on Sciences and Social Sciences, Mahasarakham, Thailand, 2016.

[25] H. A. Son, W. Jeon, J. Kim, C. Y. Heo, H. J. Yoon, J-U. Park and T-M. Chung AI-based Localization and Classification of Skin Disease with Erythema," in Scientific Reports , vol. 11, no.1, 2021.

[26] H-H. Li, Y-H. Liao, Y-N Huang and P-J. Cheng, "Based on Machine Learning for Personalized Skin Care Products Recommendation Engine," The 2020 International Symposium on Computer, Consumer and Control, Taiwan, November 2020, pp. 460-462.

[27] K. Min, G-H. Lee and S-W. Lee, "ACNet: Mask-Aware Attention with Dynamic Context Enhancement for Robust Acne Detection," The 2021 IEEE International Conference on Systems, Man, and Cybernetics, Melbourne, Australia, 2021, pp. 2724-2729.

[28] A. K. Prodeep, R. Araf, P. Ray, Md. S. A. Ulubbi, S. N. Ananna and M. F. Mridha, "Acne and Rosacea Detection from Images using Deep CNN's EfficientNet," The 2022 International Conference on Advances in Computing, Communication and Applied Informatics, Chennai, India, 2022, pp. 1-7.

[29] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," in arXiv:1804.02767, 2018.

[30] Z. Zou, Z. Shi, Y. Guo and J. Ye, "Object Detection in 20 Years: A Survey," in arXiv:1905.05055, 2019.

[31] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in arXiv:1506.01497, 2015.

[32] P. Mookdarsanit, et al., "ThaiWrittenNet: Thai Handwritten Script Recognition Using Deep Neural Networks," in Azerbaijan Journal of High Performance Computing, vol. 3, no. 1, pp. 75-93.

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

[34] L. Mookdarsanit and P. Mookdarsanit, "Thai NLP-based Text Classification of the 21st-century Skills toward Educational Curriculum and Project Design," in International Journal of Applied Computer Technology and Information Systems, vol. 11, no. 2, pp.62-67, 2021.

[35] L. Mookdarsanit and P. Mookdarsanit, " ThaiWritableGAN: Handwriting Generation under Given Information," in International Journal of Computing and Digital Systems, vol. 10, no. 1, pp. 689-699, 2021.

[36] Shuster, S. Humeau, H. Hu, A. Bordes and J. Weston, "Engaging Image Captioning via Personality," The 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019, pp. 12508-12518.

[37] P. Mookdarsanit and L. Mookdarsanit, "Thai-IC: Thai Image Captioning based on CNN-RNN Architecture," in International Journal of Applied Computer Technology and Information Systems, vol. 10, no. 1, pp.40-45, 2020.

[38] C. Saharia, et al., "Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding," in arXiv: 2205.11487, 2022.

[39] R. Rombach, A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-Resolution Image Synthesis with Latent Diffusion Models," in arXiv:2112.10752, 2021.

[40] A. Ramesh, et al., “Zero-Shot Text-to-Image Generation,” in arXiv: 2102.12092, 2021.

[41] A. Ramesh, et al., “Hierarchical Text-Conditional Image Generation with CLIP Latents,” in arXiv: 2204.06125, 2022.

[42] O. Gafni, A. Polyak, O. Ashual, S. Sheynin, D. Parikh and Y. Taigman., “Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors,” in arXiv: 2203.13131, 2022.

[43] S. Loffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in arXiv: 1502.03167, 2015.

[44] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in arXiv: 1512.03385, 2015.

[45] P. Jiang, D. Ergu, F. Liu, Y. Cai and B. Ma, "A Review of YOLO Algorithm Developments," in Procedia Computer Science, vol. 199, pp. 1066-1073, 2022.

[46] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in arXiv: 1612.08242, 2016.

[47] “LabelImg,” [Online]. Available: https://github.com/heartexlabs/labelImg

[48] W. Lui, et al. , “SSD: Single Shot MultiBox Detector,” in arXiv:1512.02325, 2015.