

# Combining convolutional neural networks and spatial-channel “squeeze and excitation” block for multiple-label image classification

Thuvanan Borvornvitchotikarn, Thongchai Yooyativong

School of Information Technology, Mae Fah Luang University, Chiang Rai, Thailand

## Article Info

### Article history:

Received Jun 1, 2022

Revised Jan 16, 2023

Accepted Mar 10, 2023

### Keywords:

Attention mechanism

Catheters and tubes

Chest X-rays

Convolutional neural networks

Multi-label image classification

## ABSTRACT

In emergency rooms and intensive care units, catheters and tubes are used to keep critically ill patients alive. Appropriate catheter or tube insertion is crucial to avoiding serious complications. Such issues can be rectified if they are identified early. Chest X-rays are commonly used to assess catheter placement. Convolutional neural networks (CNN) have recently been found to enhance multi-label classification tasks on chest X-rays images. Furthermore, attention modules have shown the effect of enhancing spatial encoding on network feature maps. This research analyzed the experiments of each CNN model with different attention blocks. Resnet200D with batch normalization and spatial-channel squeeze and excitation block (BN+scSE) is the best architecture for multiple-label image classification on a chest X-rays dataset from National Institutes of Health Clinical Center (NIH) with multiple catheters and tubes. Then came EfficientNetB5 with BN+scSE and Inception\_v3 with spatial squeeze and channel excitation block, respectively.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Thuvanan Borvornvitchotikarn

School of Information Technology, Mae Fah Luang University

Chiang Rai, Thailand

Email: 6471501002@lamduan.mfu.ac.th

## 1. INTRODUCTION

Nowadays, catheters or tubes are used to keep critically ill patients alive in emergency departments and intensive care units [1]. Tubes have many advantages. For example, an endotracheal tube (ETT) is used to provide oxygen to the lungs in the case of respiratory issues or lung illnesses. The nasogastric tube (NGT) is useful for individuals who are unable to eat or who consume less than 60% of their daily calorie requirements [2].

Misplacements of tubes can occur in three common tubes. First, in the case of ETT misplacement, the correct ETT position is between 5 and 7 cm above the carina. The Carina position is used for estimating the position of the ETT. The result of ETT misplacement is that accidental intubation of a bronchus is more common on the right main bronchus and the left main bronchus. Second, a chest X-rays is used to determine the position of the NGT. Correct the tip of the NGT position so that it is visible at least 10 cm. beyond the esophagus junction. NGT misplacement that might be passed into the airways. Tube misplacement has many issues, such as looped tube misplacement in the right main bronchus. Finally, a chest X-rays determines the central venous catheter's tip position. Correct central venous catheter (CVC) position is the position of the tip of the CVC within the superior vena cava at or just above the level of the carina. CVC misplacement could lead to many complications, such as pneumothorax or pneumomediastinum [3]. Evidently, the proportions of mispositioned cases encountered have been found to be 5%-28% of ETT, about 15% of NGT, and 2%-7% of CVC, respectively. If catheters are placed incorrectly, severe complications might arise. Commonly, chest X-rays are used to estimate the positioning of various medical equipment and to assess any associated issues

following placement or misplacement [1]. However, each day, radiologists may obtain many patients to radiograph and interpret them, and there may be delays. An assessment of tube position is complicated. Computer-aided techniques have the potential to help radiologists prioritize radiographs with possibly mispositioned catheters for interpretation and to automatically include information suggesting catheter placement in radiology reports, therefore increasing radiologists' performance [1].

In the literature, the classification of chest X-rays images of catheters and tubes has been researched for a long time. The classification methods for tubes have been proposed using various methods, including rule-based [4], and decision trees [5]. Other classification methods based on deep learning (DL) such as, Alexnet [6], GoogLeNet [6], UNET [7], ResNet [8], and EfficientNet [9]. However, these methods are still not optimal in multiple classification tasks, these methods are still inefficient because image classification involves multiple tubes from the current intubation depending upon the patient.

Recent studies on attention mechanisms show that this allows neural networks to change the informative input features, allowing them to maximize network ability and improve model performance. As a result, it has been widely used in neural networks for various machine learning problems in recent years, including, image classification [10]–[12], image segmentation [13], [14], image recognition [15], [16], and natural language processing (NLP) [17], [18]. Attention mechanisms have shown the advantages of enhancing spatial or channel encoding, e.g., squeeze-and-excitation block (SE) [10], bottleneck attention module (BAM) [19], convolutional block attention module (CBAM) [20], spatial squeeze and channel excitation block (cSE) [13], [14], channel squeeze and spatial excitation block (sSE) [13], [14], and spatial and channel squeeze and excitation block (scSE) [13], [14].

In this work, we conduct a study that focuses on the multi-label classification task. We have 11 targets in chest X-rays images, and they can be divided into 4 groups: ETT, NGT, CVC, and Swan. This paper is organized as follows. Section 2 presents the research method. Section 3 describes the experimental results and conclusion will be discussed in section 4, respectively.

## 2. METHOD

In this section, we will explain this research model for tube image classification, beginning with the implementation and moving on to the attention mechanism and the head layers. Section 2.1 presents the descriptions of the materials in our work. The details about the attention mechanism are presented in section 2.2. The explanation of well-known modules is presented in section 2.3.

### 2.1. Implementation

All the experiments were trained and tested on Processor Intel (R) Core (TM) i7-2600 CPU @ 3.40 GHz, 3,401 MHz, 4 Core(s), 8 Logical Processor(s), RAM 12.0 GB, with GPU NVidia GeForce RTX3060 12 GB. We utilized the RANCZR-CLiP dataset from the national institutes of health clinical center (NIH), which is available on Kaggle and comprises adult chest X-rays images. The RANCZR-CLiP dataset comprises 30,083 manually labeled images: 24,062 images for the training set, 6,021 images for the validation set, and 3,255 images for the test set [21]. All image data obtained in this experiment are stored in  $3 \times 448 \times 448$  pixels as shown in Figure 1. The number of each tube type shows in Figure 2. To evaluate our experiments on multiple tubes image classification task, we choose the three types of the most popular convolutional neural networks (CNN) as backbones to conduct this experiment, including Inception\_v3 [22], Resnet200D [23], and EfficientNetB5 [24]. Figure 3 shows the placing of attention module in the model of multiple-label image classification.



Figure 1. Examples of chest X-rays images with catheters and tubes [21]

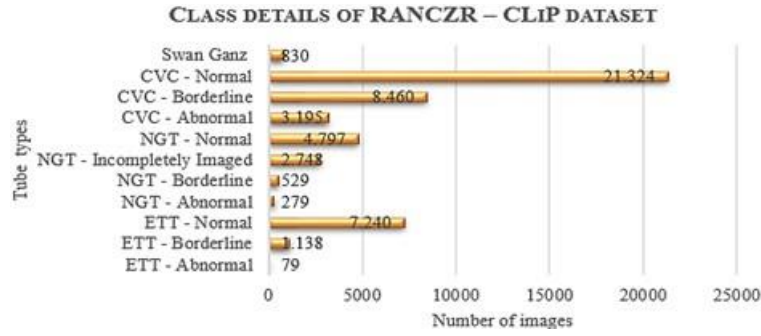


Figure 2. Number of each tube type in the dataset

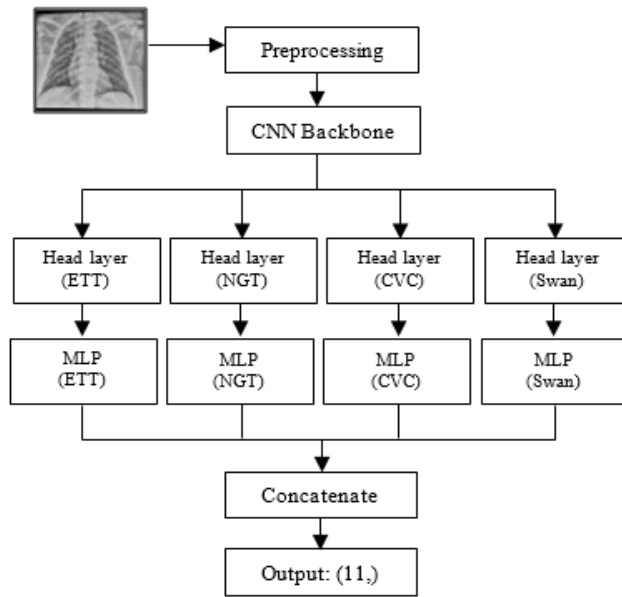


Figure 3. Multiple label image classification model MLP: (Linear -> Rectified linear units (ReLU) -> Dropout -> Linear)

**2.2. Attention mechanism**

In this section, we will briefly introduce the concept of attention mechanism through the squeeze and excitation block (SE block). SE block was proposed by Hu *et al.* [10], which was an approach for enhancing the channel relationship. The authors designed the new network to increase a network's representational ability by explicitly modeling the interdependencies between its convolutional features' channels. This block could offer a feature recalibration technique that allows the network to learn to use global information to selectively emphasize informative features and suppress less helpful features. A computational structure of the SE block is presented in Figure 1. For any transformation  $T_{tr} = F \rightarrow O$ , an input feature  $F \in R^{C' \times H' \times W'}$  and an output feature map  $O \in R^{C \times H \times W}$ .  $H$  and  $W$  represent the spatial height and width,  $C'$  and  $C$  are the input and output channels, respectively. Firstly, the input feature  $F$  are passed through a squeeze block. Let  $K = [k_1, k_2, \dots, k_c]$  is set of a 2D spatial kernel and  $\times$  represents a convolutional operator. The  $T_{tr}$  provides the output feature map as  $O = [o_1, o, \dots, o_c]$ , which is calculated using (1).

$$o_c = k_c * F = \sum_{n=1}^{c'} k_c^n * f^n, n = 1, \dots, n \tag{1}$$

After that, the transformation output  $O$  is passed to squeeze global spatial information by using global average pooling. This step can provide the necessary contextual information. Here, the  $c - th$  element of vector  $S$  is channel-wise statistics,  $i, j$  denotes spatial location. It is calculated using (2).

$$s_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W o_c(i, j), c = 1, \dots, c \tag{2}$$

Secondary, the squeeze block is followed by an excitation block that tries to completely capture channel-wise dependencies to make use of the information accumulated during the squeeze process. To achieve this goal, these must learn a non-mutually exclusive connection. This second step utilizes the sigmoid activation, which concerns the enable several channels to be highlighted rather than one-hot activation. To reduce the complexity, two fully connected (FC) layers are used to reduce a dimension  $D_1$  and adopted ReLU and FC layer  $D_2$  to increase the dimension. Where  $r$  denotes the reduction ratio (here:  $r = 16$ ),  $D_1 \in \mathbb{R}^{\frac{c}{r} \times c}$ , and  $D_2 \in \mathbb{R}^{c \times \frac{c}{r}}$ ,  $\delta$  represents the ReLU function,  $\sigma$  is Sigmoid function. The final output  $\tilde{O}$  of SE block is recalled using the multiply by  $F$ , which is calculated using (4).

$$Z = \sigma(D_2 \delta(D_1 S)) \quad (3)$$

$$\tilde{O} = z_c \cdot f_c \quad (4)$$

### 2.3. Head layers

This work will focus on techniques of enhancing the performance of CNN in the context of multiple-label image classification. In the following subsections, we will briefly explain well-known and widely used modules, i.e., SA [25], pre-activation-spatial attention (Pac-SA) [26], cSE [10], sSE [13], batch normalization (BN)+sSE, scSE [13], [14], and BN+ scSE blocks. In our experiment, all these approaches will be used as benchmarks against our method.

#### 2.3.1. Spatial attention block

A spatial-wise attention block ignores the channel information and equally preserves the feature of different channels.  $c^{3 \times 3,1}$  refers to a  $3 \times 3$  convolution with the one channel number of the hidden feature map,  $c^{3 \times 3,m}$  refers to a  $3 \times 3$  convolution with the  $m$  channel number of hidden feature map and  $m$  is integer number of output channel number needed, BN denotes the batch normalization operation [25]. In our work, we used  $m$  those values [64, 32, 8] respectively. The spatial weights map is adaptively generated by this module to properly represent the spatial connection. Spatial attention can be computed as in (5).

$$\tilde{O}_{SA} = F \times \sigma(\left(\text{BN}(c^{3 \times 3,1} \delta(\text{BN}(c^{3 \times 3,m}(F))))\right)) \quad (5)$$

#### 2.3.2. Pre-activation-spatial attention block

To reduce the parameter number of network and increase the performance of multiple-label image classification, we introduce to used spatial attention and pre-activation residual network, namely a pre-activation-spatial attention (Pac-SA) block. The Pac-SA block has three BN activations, two ReLU activations, one  $3 \times 3$  convolution with eight channel number of the hidden feature map, one  $3 \times 3$  convolution with one channel number of the hidden feature map, and one Sigmoid activation. It places BN and ReLU as pre-activation units before convolution operation. The recalibrated values are multiplied by the feature map's different weights to get the output in (6).

$$\tilde{O}_{PAC-SA} = F \times \sigma(\left(\text{BN}(c^{3 \times 3,1} \delta(\text{BN}(c^{3 \times 3,8} \delta(\text{BN}(F))))\right)) \quad (6)$$

#### 2.3.3. Spatial squeeze and channel excitation block

Roy *et al.* [14] considered the input feature map  $f^n$ , cSE block is used to squeeze the spatial information of  $f^n$  that is passed by (1) followed by global average pooling in (2). To excite the channel information of the squeezed vector  $S$ , these can be expressed in (3). The final step shows that the importance of  $\tilde{O}_{cSE}$  which is calculated using (7).

$$\tilde{O}_{cSE} = [\sigma(s_1)f_1, \dots, \sigma(s_1)f_c] \quad (7)$$

#### 2.3.4. Channel squeeze and spatial excitation block

Roy *et al.* [13] introduced sSE, which compress the information from all channels by convolution  $c^{1 \times 1}$ , the feature maps of  $F$  are squeezed along the channel and are excited the spatial data. Let  $F = [f_{1,1}, f_{1,2}, \dots, f_{i,j}, \dots, f_{H,W}]$  represent the slice on the spatial dimension, where spatial location  $(i, j)$ . In this experiment, we adopted the original sSE as shown in (8) and applied BN after the sigmoid function as shown in (9) for this experiment.

$$\tilde{O}_{sSE} = F \times \sigma(c^{1 \times 1}(F)) \quad (8)$$

$$\tilde{O}_{BN+sSE} = F \times \sigma(BN(c^{1 \times 1}(F))) \quad (9)$$

### 2.3.5. Spatial and channel squeeze and excitation block

Roy *et al.* [14] presented scSE, which declared cSE see in (7) and sSE see in (8) blocks in parallel and both outputs are added in (10). scSE could generate both channel re-scaling and spatial re-scaling. The feature maps are recalibrated to learn more meaningful feature maps, which are provided both in terms of spatial and channel relevance see in (10). For improving the performance of CNN model and attention block, we proposed BN that normalize the input feature maps and multiplied the output of cSE and sSE blocks. This could get both spatial and channel importance of each location on feature maps cSE and sSE elementwise see in (11).

$$\tilde{O}_{scSE} = \tilde{O}_{cSE} + \tilde{O}_{sSE} \quad (10)$$

$$\tilde{O}_{BN+scSE} = \tilde{O}_{cSE} \times \tilde{O}_{BNsSE} \quad (11)$$

## 3. RESULTS AND DISCUSSION

This section will discuss the outcomes of each experiment. Inception\_v3 [22], Resnet200D [23], and EfficientNetB5 [24] architecture models were tested on training and validation data. In addition, our experiment should provide a concise and pre-cise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn in comparison with other modules, initial baseline SA [25], PAC-SA [26], cSE [10], sSE [13], BN+sSE, scSE [13], [14], and BN+ scSE blocks. The goal of this experiment is to determine the mean of accuracy score (mAs) and mean of average precision score (mAPs) of each architectural model during the image training process. This experiment will be run using 16 training epochs. Tables 1 to 3 presents the number of network parameters, mAs, and mAPs of each CNN architecture with different attention layers.

As computes the accuracy of the model, the function returns 1, 0 if the subset accuracy is 1, if the entire set of predicted labels matches the true set of labels, otherwise it is 0. In this experiment, we trained CNN models in 5-fold validations  $v$ , where  $y'_i$  denotes the predicted value of the  $i$ -th sample,  $y_i$  represents is the corresponding true value, and the fraction of correct predictions overs. The mean of  $As$  can evaluate in (13).

$$As = \frac{1}{s} \sum_{i=0}^{s-1} 1(y'_i = y_i) \quad (12)$$

$$mAs = \frac{1}{v} \sum_{v=1}^{a=v} As_a \quad (13)$$

In  $APs$  computation, the weighted mean of precision acquired at each threshold is used to summarize a precision-recall curve, with the increase in recall from the previous threshold used as the weight.  $R_n$  and  $P_n$  denote the recall and precision at the  $n_{th}$  threshold. The mean of  $APs$  can calculate in (15).

$$APs = \sum_n (R_n - R_{n-1}) P_n \quad (14)$$

$$mAPs = \frac{1}{v} \sum_{v=1}^{a=v} APs_a \quad (15)$$

Table 1 presents the results of the Inception\_v3 with different attention blocks. cSE block got the best results from all the attention blocks, with  $mAs$  94.68%,  $mAPs$  65.34%. Then, followed by the three high performances, including the scSE got  $mAs$  94.70%,  $mAPs$  65.24%, the SA got  $mAs$  94.73%,  $mAPs$  64.96%, and the BN+sSE got  $mAs$  94.68%,  $mAPs$  66.01%, respectively.

Table 1. Number of network parameters, mean of accuracy score, and mean of average precision score of Inception\_v3 with different attention blocks

Model	Attention module	Parameters	$mAs$	$mAPs$
Inception_v3 [22]	baseline	20,227,637	93.76	62.49
	cSE [10]	40,699,371	94.68	65.34
	SA [25]	43,413,939	94.73	64.96
	PAC-SA [26]	43,431,219	94.28	63.14
	sSE [13]	38,601,715	94.66	64.77
	BN+ sSE	40,707,563	94.68	66.01
	scSE [13], [14]	40,707,563	94.70	65.24
	BN+ scSE	40,707,571	<b>93.22</b>	<b>61.48</b>

Table 2 presents the results of the Resnet200D with different attention blocks. BN+ scSE block got the best results from all the attention blocks, with  $mAs$  95.87%,  $mAPs$  72.15%. Then, followed by the three high performances, including the scSE got  $mAs$  95.85%,  $mAPs$  71.31%, the BN+sSE got  $mAs$  95.79%,  $mAPs$  70.46%, and SA got  $mAs$  95.79%,  $mAPs$  70.42%, respectively.

Table 3 presents the results of the EfficientNetB5 with different attention blocks. BN+ scSE block got the best results from all the attention blocks, with  $mAs$  95.24%,  $mAPs$  68.44%. Then, followed by the three high performances, including the scSE got  $mAs$  95.24%,  $mAPs$  68.21%, the BN+sSE got  $mAs$  95.21%,  $mAPs$  68.07%, and PAC-SA got  $mAs$  95.16%,  $mAPs$  68.02%, respectively.

Table 2. Number of network parameters, mean of accuracy score, and mean of average precision score of Resnet200D with different attention blocks

Model	Attention module	Parameters	$mAs$	$mAPs$
Resnet200D [23]	baseline	66,862,955	94.31	63.30
	cSE [10]	81,557,867	95.81	69.86
	SA [25]	84,272,435	95.79	70.42
	PAC-SA [26]	84,289,715	95.77	70.32
	sSE [13]	79,460,211	95.72	69.57
	BN+ sSE	79,460,211	95.79	70.46
	scSE [13], [14]	81,566,059	95.85	71.31
	BN+ scSE	81,566,059	<b>95.87</b>	<b>72.15</b>

Table 3. Number of network parameters, mean of accuracy score, and mean of average precision score of EfficientNetB5 with different attention blocks

Model	Attention module	Parameters	$mAs$	$mAPs$
EfficientNetB5 [24]	baseline	32,559,675	94.33	63.72
	cSE [10]	47,254,587	95.17	67.09
	SA [25]	49,969,155	94.51	65.29
	PAC-SA [26]	49,986,435	95.16	68.02
	sSE [13]	45,156,931	94.98	66.72
	BN+ sSE	45,156,931	95.21	68.07
	scSE [13], [14]	47,262,779	95.24	68.21
	BN+ scSE	47,262,787	<b>95.24</b>	<b>68.44</b>

According to these findings, overall, Resnet200D with the addition of the Bn+scSE block outperforms all trained CNN architectural models by  $mAs$  95.87% and  $mAPs$  72.15% see details in Table 2. The EfficientNetB5 architecture followed, with  $mAs$  95.24% and  $mAPs$  68.44% see details in Table 3, and Inception\_v3 with  $mAs$  94.68%,  $mAPs$  65.34%. See details in Table 1, respectively.

#### 4. CONCLUSION

This paper has evaluated the experiments of each CNN model with different attention blocks. In order by CNN model, Resnet200D with the improved scSE is the best architecture for multiple-label image classification on a chest X-rays dataset with multiple catheters and tubes from NIH. Then came EfficientNetB5 with improved scSE and Inception\_v3 with cSE. This experiment shows the higher performance of the CNN model with the attention block than the baseline CNN model, which could improve the terms of  $mAs$ , and  $mAPs$ . In addition, we have improved the attention module, i.e., the improved sSE and the improved scSE. They use BN to normalize the input feature maps of these attention blocks, which can reduce covariate shift and reduce the effects of exploding and vanishing gradients.

#### ACKNOWLEDGEMENTS

This research is part of a PhD dissertation, and the authors would like to thank Mae Fah Luang University for funding the publication of this research.

#### REFERENCES

- [1] X. Yi, S. J. Adams, R. D. E. Henderson, and P. Babyn, "Computer-aided assessment of catheters and tubes on radiographs: how good is artificial intelligence for assessment?," *Radiology: Artificial Intelligence*, vol. 2, no. 1, Jan. 2020, doi: 10.1148/ryai.2020190082.
- [2] N. D. P. Concepcion, B. F. Laya, and E. Y. Lee, "Current updates in catheters, tubes and drains in the pediatric chest: a practical evaluation approach," *European Journal of Radiology*, vol. 95, pp. 409–417, Oct. 2017, doi: 10.1016/j.ejrad.2016.06.015.
- [3] L. R. Goodman, P. A. Conrardy, F. Laing, and M. M. Singer, "Radiographic evaluation of endotracheal tube position," *American Journal of Roentgenology*, vol. 127, no. 3, pp. 433–434, Sep. 1976, doi: 10.2214/ajr.127.3.433.




- [4] E.-F. Kao, T.-S. Jaw, C.-W. Li, M.-C. Chou, and G.-C. Liu, "Automated detection of endotracheal tubes in paediatric chest radiographs," *Computer Methods and Programs in Biomedicine*, vol. 118, no. 1, pp. 1–10, Jan. 2015, doi: 10.1016/j.cmpb.2014.10.009.
- [5] D. Lederman, S. Lampotang, and M. Y. Shamir, "Automatic endotracheal tube position confirmation system based on image classification—a preliminary assessment," *Medical Engineering & Physics*, vol. 33, no. 8, pp. 1017–1026, Oct. 2011, doi: 10.1016/j.medengphy.2011.04.006.
- [6] V. Singh, V. Danda, R. Gorniak, A. Flanders, and P. Lakhani, "Assessment of critical feeding tube malpositions on radiographs using deep learning," *Journal of Digital Imaging*, vol. 32, no. 4, pp. 651–655, May 2019, doi: 10.1007/s10278-019-00229-9.
- [7] X. Yi, S. Adams, P. Babyn, and A. Elnajmi, "Automatic catheter and tube detection in pediatric X-ray images using a scale-recurrent network and synthetic data," *Journal of Digital Imaging*, vol. 33, no. 1, pp. 181–190, Apr. 2019, doi: 10.1007/s10278-019-00201-7.
- [8] R. D. E. Henderson, X. Yi, S. J. Adams, and P. Babyn, "Automatic detection and classification of multiple catheters in neonatal radiographs with deep learning," *Journal of Digital Imaging*, vol. 34, no. 4, pp. 888–897, Jun. 2021, doi: 10.1007/s10278-021-00473-y.
- [9] A. B. M. Khan and S. M. A. Ali, "Early detection of malpositioned catheters and lines on chest X-rays using deep learning," Jun. 2021, doi: 10.1109/icaicst53116.2021.9497809.
- [10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," Jun. 2018, doi: 10.1109/cvpr.2018.00745.
- [11] D. Xin, Y.-W. Chen, and J. Li, "Fine-grained butterfly classification in ecological images using squeeze-and-excitation and spatial attention modules," *Applied Sciences*, vol. 10, no. 5, p. 1681, Mar. 2020, doi: 10.3390/app10051681.
- [12] Y. Han, C. Wei, R. Zhou, Z. Hong, Y. Zhang, and S. Yang, "Combining 3D-CNN and squeeze-and-excitation networks for remote sensing sea ice image classification," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–15, Apr. 2020, doi: 10.1155/2020/8065396.
- [13] A. G. Roy, N. Navab, and C. Wachinger, "Recalibrating fully convolutional networks with spatial and channel 'squeeze and excitation' blocks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 540–549, Feb. 2019, doi: 10.1109/tmi.2018.2867261.
- [14] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2018*, Springer International Publishing, 2018, pp. 421–429.
- [15] J. Fu, H. Zheng, and T. Mei, "Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition," Jul. 2017, doi: 10.1109/cvpr.2017.476.
- [16] J. Ba, V. Mnih, and K. Kavukcuoglu, "Multiple object recognition with Visual attention," Dec. 2014.
- [17] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," *Advances in Neural Information processing systems*, vol. 27, Jun. 2014.
- [18] A. Vaswani *et al.*, "Attention is all you need," in *Proceedings of the 31st Conference on Neural Information Processing Systems*, Dec. 2017, pp. 5998–6008, doi: 10.48550/arXiv.1706.03762.
- [19] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "BAM: bottleneck attention module," Jul. 2018.
- [20] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2\_1.
- [21] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," Jul. 2017, doi: 10.1109/cvpr.2017.369.
- [22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2016, vol. 2016-December, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908 LNCS, pp. 630–645, 2016, doi: 10.1007/978-3-319-46493-0\_38.
- [24] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," May 2019.
- [25] P. Zhao, J. Zhang, W. Fang, and S. Deng, "SCAU-Net: spatial-channel attention U-Net for gland segmentation," *Frontiers in Bioengineering and Biotechnology*, vol. 8, Jul. 2020, doi: 10.3389/fbioe.2020.00670.
- [26] T. Borvornvitchotikarn and T. Yooyativong, "Pre-activation-spatial attention module for multiple catheters and tubes classification," Jan. 2022, doi: 10.1109/ectidamtncon53731.2022.9720402.

## BIOGRAPHIES OF AUTHORS



**Thuvanan Borvornvitchotikarn**    received B.S. in Industrial Technology from Rajabhat Institute Chiang rai, Chiang rai, Thailand, in 2001, and M.Sc. in Internet and Information Technology from Naresuan University, Thailand, in 2006. Now, he is studying for his Ph.D. in Computer engineering at Mae Fah Luang University, Chiang Rai, Thailand. His main research interests are medical image registration, medical image classification, image processing and AI Technologies. He can be contacted at email: 6471501002@lamduan.mfu.ac.th.



**Asst.Prof.Gp.Capt.Dr. Thongchai Yooyativong**    received B.S. in Electrical Engineering from Norwich University, USA, and M.Sc. in Electrical Engineering from Colorado State University, USA and Ph.D. in Electrical Engineering from Colorado State University, USA. In Computer engineering at Mae Fah Luang University, Chiang Rai, Thailand. His main research interests are Agricultural Engineering, Data Mining, Artificial Neural Network, and Computer Communications (Networks). He can be contacted at email: thongchai.mfu@gmail.com.