Human activity recognition with self-attention

Yi-Fei Tan, Soon-Chang Poh, Chee-Pun Ooi, Wooi-Haw Tan
Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia

Article Info

Article history:

Received Mar 31, 2022
Revised Sep 15, 2022
Accepted Oct 13, 2022

Keywords:

Convolution neural network
Human activity recognition
Long short-term memory
Self-attention

ABSTRACT

In this paper, a self-attention based neural network architecture to address human activity recognition is proposed. The dataset used was collected using smartphone. The contribution of this paper is using a multi-layer multi-head self-attention neural network architecture for human activity recognition and compared to two strong baseline architectures, which are convolutional neural network (CNN) and long-short term network (LSTM). The dropout rate, positional encoding and scaling factor are also been investigated to find the best model. The results show that proposed model achieves a test accuracy of 91.75%, which is a comparable result when compared to both the baseline models.

Corresponding Author:

Yi-Fei Tan
Faculty of Engineering, Multimedia University
63100 Cyberjaya, Selangor, Malaysia
Email: yftan@mmu.edu.my

1. INTRODUCTION

Around the world, the population of elderly is increasing rapidly. According to United Nations analysis of recent trends of elderly population, the population is predicted to increase to 2.1 billion by 2050 [1]. Department of Statistics Malaysia projected the elderly population to reach about 20% of total population [2]. Therefore, there are research which focused on improving elderly via assisted living technology [3]. One of the key technologies is activity recognition. Activity recognition is a task of recognizing human activities based on a series of observations of human actions [4]. Computer scientists addressed this task by using various machine learning algorithms which feed on the observational data of human actions and classify the given data into specific activity type. The observational data of activities can be collected by using different means such as sensors or cameras [5]–[7]. In this paper, the focus is on sensor-based activity recognition algorithm because sensory-based system is much less intrusive and lightweight when compared to vision-based methods. On-body sensors such as accelerometer and gyroscope can be used to collect acceleration and angular velocity of human body in various axis [8]–[11]. On-body sensors as data collectors are ubiquitous and prevalent in human activity recognition research because of their relatively low cost. For example, electronic devices such as smartphones, smart watches and wearable activity tracker have embedded accelerometer and gyroscope. Besides, they are not limited to any location because people can carry them everywhere. One of the limitations of using on-body sensors is the battery issue. The pattern of the time series sensory data for each type of activity is distinct. For example, time series acceleration of running should have a higher rate of change compared to walking. This unique distinction in pattern allows machine learning algorithm to classify them.

The epplication of neural network architecture in human activity recognition has grown in recent years. Besides neural networks, there are several machine learning algorithms that were used for human activity recognition which include support vector machine (SVM) [12] and random forest (RF) [13]. Anguita *et al.* [12] used a dataset collected using smartphone's inertial sensors. In this research, SVM which exploits fixed-point arithmetic was used as a multiclass classifier. Contrary to traditional SVM, this approach demonstrated a considerable improvement in computational cost while achieving similar accuracy. Xu *et al.* [13] proposed a RF model which classify human activity based on sensory data from wearable devices. RF model managed to obtain an overall accuracy of 90%.

The prevalent neural network architecture includes CNN and recurrent neural network (RNN). Wu and Zhang [14] collected a 128-dimensional time series sensory data from accelerometer and gyroscope of a smartphone. In this study, they trained an 8-layer CNN network to classify the sensory data into type of activities. Their method achieved comparable results with conventional and state of the art models. Xi *et al.* [15] reasoned using pooling after convolution to expand the receptive fields is not advisable because it could bring about information loss. They proposed an alternative model based on dilated convolution which can expand receptive field without compromising on information loss.

Long-short term network (LSTM) [16] is a widely used variant of RNN. Güney and Erdas [17] used LSTM to classify raw sensory data collected from three-axis accelerometer. The proposed method managed to obtain a classification accuracy of 91.34%. Mahmud *et al.* [18] used a multi-stage LSTM to extract temporal features from sensory data and achieved a F1 score of 83.9%. Each stage of LSTM is used to extract features from each sensor's data. Then, the output of every stage of LSTM are aggregated using fully connected layers.

There are also research works which combine both CNN and RNN for human activity recognition. Mutegeki and Han [19] introduced a hybrid model of CNN and LSTM which achieved high classification accuracy and reduced the complexity of the model. In [20], a hybrid model based on CNN and gated recurrent unit (GRU) is proposed. GRU is another variant of RNN. In this work, the authors trained the model on recognition of pairwise similar activities. This means that this model can be trained on data of one type of activity and classify the data of another similar activity type.

In this work, we developed a human activity classification algorithm based on multi-head self-attention neural network architecture. We trained and evaluated our models on a dataset which consists of six activities. The proposed model was compared with two strong baselines based on CNN and LSTM respectively. This work consists of ablation study of various component that is commonly used in self-attention model such as dropout, scaling study and positional encoding. Our results show that the proposed self-attention neural network architecture delivered a test accuracy on unseen data which is on-par with both the baseline architectures. The organization of this paper is: in section 2 descriptions of the dataset are given; in section 3 we present the detailed architecture of baseline CNN and LSTM model, as well as proposed self-attention model; section 4, the experimental result and discussion are discussed; section 5, we conclude the paper with a conclusion and future work.

2. DATASET

The dataset used in this research is a publicly available human activity recognition dataset called UCI-HAR dataset [21]. This data is collected from 30 person aged 19 to 48 years old. Each volunteer was asked to carry out six types of activities (walking, walking upstairs, walking downstairs, standing, sitting, and laying) wearing a Samsung Galaxy smartphone on the waist. The dataset is not raw sensory data of tri-axis accelerometer and gyroscope. It was pre-processed with noise filters and sampled with a sliding window of 2.56 seconds with 50% overlap. Eventually, each instance of sensory data consists of 128 readings or time steps. The dataset consists of 10,299 instances in total. The authors of the dataset split the dataset into 70% training set and 30% test set based on the volunteers. This means that volunteers for training set are not included in test set and vice versa. The percentage of how much each activity type made up the train and test set and other details of the dataset is summarized in Table 1.

Table 1. Class percentage for train and test set

Activity Type	Train (%)	Train (%)
Walking	16.68	16.83
Walking upstairs	14.59	15.98
Walking downstairs	13.41	14.25
Sitting	17.49	16.66
Standing	18.69	18.05
Laying	19.14	18.22
Total number of instances	7352	2947

There are nine features available for each instance of data which include triaxial acceleration from the accelerometer (total acceleration), the estimated body acceleration and the triaxial angular velocity from the gyroscope. Each of these three categories contribute 3 features in *x*, *y* and *z* axis. Each instance of data consists of 128 time steps and 9 dimensions per time step for each of the features and is a matrix of size 128×9.

3. RESEARCH METHOD

Time series classification problem is usually solved using certain type of RNN architecture such as LSTM and GRU. However, RNN computation cannot be done in parallel for all the time steps. RNN architecture design requires computation to be computed sequentially. Due to this fact, CNN is explored as an alternative for time series data because CNN can compute all the time steps in parallel with convolution. In this work, we developed two baselines, one based on LSTM and another based on CNN. We compared our multi-head self-attention model to these baselines. Self-attention models are widely used model in natural language processing (NLP) tasks which are also sequential in nature.

3.1. CNN architecture

The CNN baseline model has 2 layers of 1D convolutional layer with filter of size 3. The number of filters for both layers are 64. Both of the convolutional layers use non-linear activation function called rectified linear unit (ReLU). The output of second convolutional layer is passed to a 1D max pooling layer with pooling size of 2. The output is then flattened and connected to two layers of fully connected layers. The first fully-connected layer has 128 hidden units and uses ReLU activation function. The final layer has 6 units corresponding to 6 classes of activities and uses a SoftMax activation function. Figure 1 visualized the CNN model which serves as one of our baseline models.

3.2. LSTM architecture

The LSTM baseline model has 3 layers of LSTM. The first LSTM layer has 32 hidden units. The second LSTM layer has 64 hidden units. The third LSTM layer has 32 hidden units. The final time step of the third LSTM layer is connected to a fully connected layer. The fully connected layer has 6 units corresponding to 6 classes of activities and uses a SoftMax activation function. Figure 2 summarizes the LSTM model we used.

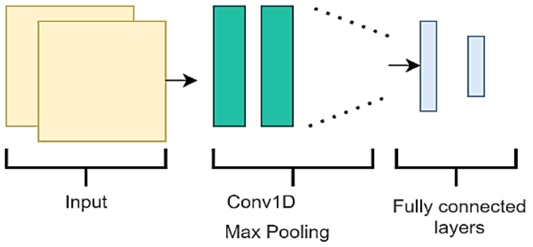

Figure 1. CNN model

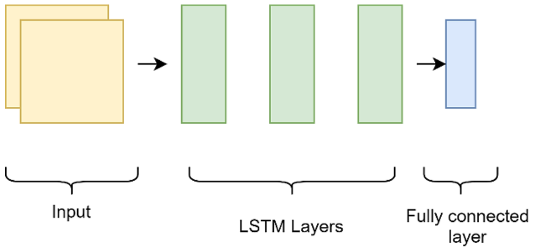

Figure 2. LSTM model

3.3. Multi-head self-attention architecture

3.3.1. Multi-head self-attention layer

Recently, self-attention neural network architecture [22] is the go-to neural network architecture for many NLP tasks. Large language models with lots of self-attention layers such as bidirectional encoder representations from transformers (BERT) can be trained on large text corpus without annotations [23]. Then, it acts a pre-trained model for downstream NLP tasks such as question answering (QA) [24]. Recently, researchers are investigating the possibility of using attention-based model for computer vision tasks such as image classification [25] and object detection [26].

In [22], Google scientists introduced fully attention-based sequence-to-sequence model called transformer. Transformer consists of a basic building block of self-attention called scaled dot product attention which is given by equations listed at (1). The input sequence, X is a tensor of N×T×F where N is the number of data examples, T is the number of time steps and F is the number of features. For UCI-HAR dataset, T=128 and F=9. The input X is passed through three fully connected layer to produce Q, K ad V. All three fully connected layers are linear without any non-linear activation function. Their parameters include the weights denoted as W_q, W_k, W_v and biases denoted as b_q, b_k and b_v. Each of the weight term is a matrix of F×H where H is the number of hidden units of the fully connected layer. In addition, each of the bias term is a vector of H dimension. These fully connected layers produce three output tensors denoted by Q, K and V.

Each of these output tensors is a tensor of $N×T×H$. We then carry out dot product for both Q and K with a scaling factor of $\frac{1}{\sqrt{d_k}}$.

$$Q = XW_q + b_q$$
$$K = XW_k + b_k$$
$$V = XW_v + b_v$$
$$\text{Attention}(Q,K,V)=\text{SoftMax}\ (\frac{Q \cdot K^T}{\sqrt{d_k}})V \tag{1}$$

The output is then passed through a SoftMax activation and multiplied with V which produce a final tensor of $N×T×H$.

3.3.2. Multi-head self-attention model

Figure 3 shows the multi-head self-attention model for human activity recognition for one data instance. A data instance of 128×9 is expanded in dimension to 128×128 after passing through a fully connected layer (labelled FC in Figure 3) with ReLU activation. Then, the matrix of 128×128 is passed to 2 blocks of "4-head self-attention block". Each 4-head self-attention block consists of 2 layers. The first layer is a 4-head self-attention layer described in Part 1 of this subsection. Each individual self-attention layer outputs a 128×128 matrix. The 4-head self-attention layer results in four 128×128 matrices which is then concatenated to a matrix of 128×512. The output matrix of this concatenated matrix is passed to a fully connected layer with 128 hidden units, which then outputs a 128×128 matrix. The output matrix of 2 blocks of self-attention is matrix of 128×128, which is then unrolled to a column vector of 16384 dimensions. The column vector is then passed through a FC layer with 128 units using ReLU and then to the prediction layer with six units with SoftMax activation. The output is a vector of 6. Finally, we added dropout for some FC layer. Table 2 summarizes the model and listed the number of parameters for each layer.

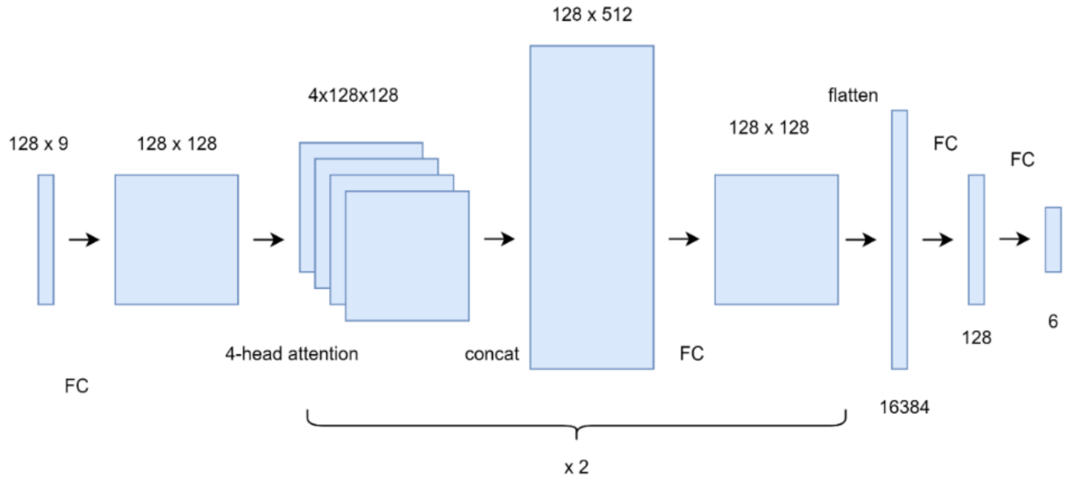

Figure 3. Multi-head self-attention model used for time series activity recognition

Table 2. Model summary

Layer	Output Shape	Number of parameters
FC (128 units, ReLU)	128×128	1280
4-head attention	4×128×128	4×49,536
Concatenate	128×512	-
FC (128 units, ReLU, dropout)	128×128	65,664
4-head attention	4×128×128	4×49,536
Concatenate	128×512	-
FC (128 units, ReLU, dropout)	128×128	65,664
Flatten	16384	-
FC (128 units, ReLU, dropout)	128	2,097,280
FC (6 units, Softmax)	6	774
FC (128 units, ReLU)	128×128	1,280
Total	-	2,626,950

3.3.3. Positional encoding

Self-attention layer does not have any convolution and recurrence. For an instance of data, LSTM takes in each time step of the data recurrently. As for CNN, each position of the output feature map of convolutional layer corresponds to a receptive field in the input. Both LSTM and CNN have location information of each time step of the sequential input. On the contrary, the output at each time step of the self-attention layer is independent of computation at other time steps. This results in self-attention is not location or positionally aware. For example, if the input is the word “apple”. To self-attention, “apple” or its anagrams such as “papel” are the same. Therefore, the authors proposed that it is vital to inject some positional information to the input.

In this work, we tried the positional encoding function introduced in [22]. The positional encoding function generates a matrix PE of same shape as input X. The positional information is injected using elementwise addition of X and PE as given in (2).

$$X = X + PE \tag{2}$$

4. RESULTS AND DISCUSSION

We evaluated the proposed self-attention on test set using the evaluation metric, accuracy. For every experiment, we used Adam optimizer and run the training for 50 epochs. We picked the best test accuracy out of all the 50 epochs to be presented.

4.1. Dropout

Dropout is a regularization method [27]. We added dropout at some of the FC layer. The detailed position of dropout in the self-attention model is given in Table 2. Since the dropout rate is a hyperparameter, we have tried out a few choices of dropout rate. Table 3 listed the test accuracy of our self-attention model for difference choices of dropout rate. It was discovered that self-attention model with dropout rate of 0.2 has the highest test accuracy of 0.9175. Therefore, we fixed dropout rate to 0.2 for later experiments.

Table 3. Choice of dropout rate and test accuracy

Dropout Rate	Test Accuracy
None	0.9070
0.1	0.9152
0.2	0.9175
0.3	0.9104
0.5	0.9023

4.2. Positional encoding

Experiment was carried out to investigate positional encoding. Our experiment result is tabulated in Table 4. Addition of positional encoding results in a test accuracy of 0.8985 which is less accurate than the model without addition of positional encoding. Positional encoding does not improve the performance of our proposed self-attention model. We believed the reason behind this is due to the sequence length of 128 for this dataset is relatively shorter compared to NLP tasks. Besides, we conjecture that the activity recognition dataset is sensitive to noise. Addition of positional encoding may have brought about a regularizing side effect.

4.3. Scaling factor

In this part, we evaluated the necessity of scaling factor. Scaling factor $\frac{1}{\sqrt{d_k}}$ is part of the self-attention layer given in (1). For our proposed model, d_k=128. By setting d_k=1, the scaling factor’s effect disappear as it is multiplied by 1. We discovered adding scaling factor will hurt the accuracy of the model. Table 5 shows the effect of scaling factor on test accuracy. Addition of scaling factor results in a model with test accuracy of 0.8853 which is lower than a model without scaling factor. Therefore, the model that was used to compare with the two baselines did not include scaling factor.

Table 4. Effects of positional encoding test accuracy

With Positional Encoding	Test Accuracy
Yes	0.8985
No	**0.9175**

Table 5. Effects of scaling factor on test accuracy

d_k	Train Accuracy	Test Accuracy
1	0.9643	**0.9175**
128	0.9519	0.8853

4.4. Comparison with baseline models

In this paper, we introduced two baseline models which include CNN and LSTM. We compared the performance of our proposed self-attention model with these two baselines. Table 6 listed the train and test accuracies of self-attention model and two baseline models: CNN and LSTM. The model with the highest test accuracy is LSTM followed by CNN. Our proposed model has the lowest test accuracy of 0.9175 with less than 1% difference from LSTM's test accuracy of 0.9237. Our proposed model is comparable to both the baseline models.

Table 6. Train and test accuracy for self-attention model, CNN and LSTM

Model	Train Accuracy	Test Accuracy
CNN	0.9536	0.9223
LSTM	0.9559	**0.9237**
Self-attention	0.9643	0.9175

5. CONCLUSION AND FUTURE WORK

In this paper, an activity recognition algorithm using multi-head self-attention neural network architecture is introduced. The experiment to evaluate this proposed model was conducted using UCI-HAR dataset. The results demonstrated that the proposed model has a test accuracy of 0.9175 which is comparable to two baseline models: CNN and LSTM. In addition, we also discovered some of the common components of self-attention layer such as scaling factor and positional encoding actually reduce test accuracy. In future work, we would like to apply our model on other datasets and collect our own dataset.

REFERENCES

[1] WHO, "Ageing and health," *World Health Organization*, 2021. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (accessed Oct. 04, 2021).

[2] M. U. Mahidin, *Selected demographic indicators*. Department of Statistics Malaysia, 2018.

[3] G. Cicirelli, R. Marani, A. Petitti, A. Milella, and T. D'Orazio, "Ambient assisted living: a review of technologies, methodologies and future perspectives for healthy aging of population," *Sensors*, vol. 21, no. 10, May 2021, doi: 10.3390/s21103549.

[4] A. K. S. Kushwaha, O. Prakash, A. Khare, and M. H. Kolekar, "Rule based human activity recognition for surveillance system," in *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, Dec. 2012, pp. 1–6, doi: 10.1109/IHCI.2012.6481853.

[5] M. C. Sorkun, A. E. Danisman, and O. D. Incel, "Human activity recognition with mobile phone sensors: Impact of sensors and window size," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404569.

[6] A. Bagate and M. Shah, "Human activity recognition using RGB-D sensors," in *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, May 2019, pp. 902–905, doi: 10.1109/ICCS45141.2019.9065460.

[7] S. Tao, M. Kudo, H. Nonaka, and J. Toyama, "Camera view usage of binary infrared sensors for activity recognition," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 1759–1762.

[8] T. Kaytaran and L. Bayindir, "Activity recognition with wrist found in photon development board and accelerometer," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404630.

[9] S-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Feb. 2017, pp. 131–134, doi: 10.1109/BIGCOMP.2017.7881728.

[10] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1780–1786, Jun. 2014, doi: 10.1109/TBME.2014.2307069.

[11] N. Hardiyanti, A. Lawi, Diaraya, and F. Aziz, "Classification of human activity based on sensor accelerometer and gyroscope using ensemble SVM method," in *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, Nov. 2018, pp. 304–307, doi: 10.1109/EIConCIT.2018.8878627.

[12] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2012, pp. 216–223.

[13] L. Xu, W. Yang, Y. Cao, and Q. Li, "Human activity recognition based on random forests," in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Jul. 2017, pp. 548–553, doi: 10.1109/FSKD.2017.8393329.

[14] W. Wu and Y. Zhang, "Activity recognition from mobile phone using deep CNN," in *2019 Chinese Control Conference (CCC)*, Jul. 2019, pp. 7786–7790, doi: 10.23919/ChiCC.2019.8865142.

[15] R. Xi, M. Hou, M. Fu, H. Qu, and D. Liu, "Deep dilated convolution on multimodality time series for human activity recognition," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489540.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[17] S. Guney and C. B. Erdas, "A deep LSTM approach for activity recognition," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2019, pp. 294–297, doi: 10.1109/TSP.2019.8768815.

[18] T. Mahmud, S. S. Akash, S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Human activity recognition from multi-modal wearable sensor data using deep multi-stage LSTM architecture based on temporal feature aggregation," in *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug. 2020, pp. 249–252, doi: 10.1109/MWSCAS48704.2020.9184666.

[19] R. Mutegeki and D. S. Han, “A CNN-LSTM approach to human activity recognition,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Feb. 2020, pp. 362–366, doi: 10.1109/ICAIIC48513.2020.9065078.

[20] M. S. Siraj and M. A. R. Ahad, “A hybrid deep learning framework using CNN and GRU-based RNN for recognition of pairwise similar activities,” in *2020 Joint 9th International Conference on Informatics, Electronics and Vision (ICIEV) and 2020 4th International Conference on Imaging, Vision and Pattern Recognition (icIVPR)*, Aug. 2020, pp. 1–7, doi: 10.1109/ICIEVicIVPR48672.2020.9306630.

[21] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” in *Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning*, 2013, pp. 437–442.

[22] A. Vaswani *et al.*, “Attention is all you need,” *Advances in neural information processing systems*, Jun. 2017.

[23] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” *arXiv:1810.04805*, Oct. 2018.

[24] D. Luo, J. Su, and S. Yu, “A BERT-based approach with relation-aware attention for knowledge base question answering,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–8, doi: 10.1109/IJCNN48605.2020.9207186.

[25] A. Dosovitskiy *et al.*, “An image is worth 16x16 Words: transformers for image recognition at scale,” *arXiv:2010.11929*, Oct. 2020.

[26] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, “End-to-end object detection with transformers,” in *Computer Vision ECCV 2020*, Springer International Publishing, 2020, pp. 213–229.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

BIOGRAPHIES OF AUTHORS

Yi-Fei Tan received her B.Sc. (Hons), M.Sc. and Ph.D. from University of Malaya (UM), Malaysia. She is currently a senior lecturer at the Faculty of Engineering in Multimedia University (MMU), Cyberjaya, Malaysia. Her research areas include machine learning, deep learning, image processing, big data analytics and queueing theory. She can be contacted at email: yftan@mmu.edu.my.

Soon-Chang Poh obtained his B. Eng. (Hons.) Electronics and Master of Engineering Science from Multimedia University, Malaysia. He is currently a research officer at the Faculty of Engineering in Multimedia University (MMU), Cyberjaya, Malaysia. His research interests include deep learning and anomaly detection. He can be contacted at email: psoonchang@gmail.com.

Chee-Pun Ooi obtained his Ph.D. in Electrical Engineering from University of Malaya, Malaysia, in year 2010. His current position is an Assoc. professor the Faculty of Engineering in Multimedia University, Malaysia. He is a chartered engineer registered with Engineering Council (the British regulatory body for Engineers), member of IET. His research areas are FPGA based embedded systems and embedded systems. He can be contacted at email: cpooi@mmu.edu.my.

Wooi-Haw Tan received his M.Sc. in Electronics from Queen’s University of Belfast, UK and a Ph.D. in Engineering from Multimedia University. He is currently a senior lecturer at Multimedia University. Dr. Tan’s areas of expertise include image processing, embedded system design, internet of things (IoT), machine learning and deep learning. He can be contacted at email: twhaw@mmu.edu.my.