Micro-Doppler-Radar-Based UAV Detection Using Inception-Residual Neural Network

Hai Le[1], Van-Sang Doan[2], Dai Phong Le[1], Huu-Hung Nguyen[1], Thien Huynh-The[3], Khanh Le-Ha[1], Van-Phuc Hoang[1*]
[1]Le Quy Don Technical University, no. 236 Hoang Quoc Viet Str., Hanoi, Vietnam
[2]Vietnam Naval Academy, no. 30 Tran Phu Str., Nha Trang, Vietnam
[3]Kumoh National Institute of Technology, Gumi-si, South Korea
[*]Corresponding Author: phuchv@lqdtu.edu.vn

***Abstract*—This paper demonstrates the performance evaluation of UAV detection based on micro-Doppler radar image data with the proposed inception-residual neural network (IRNN). Accordingly, the network is designed and analyzed by changing network hyper-parameters through experiment with the Real Doppler RAD-DAR (RDRD) dataset that is collected by the practical measurements. Numerical analysis results show that the proposed network with 16 filters yield a good trade-off between accuracy and time-consuming performances. Moreover, the network is taken into account for competing with three other networks. Due to inception-residual structure, the proposed network remarkably outperforms other ones.**

***Index Terms*—Neural network, Micro-Doppler radar, Inception-residual neural network, UAV detection**

I. Introduction

Unmanned aerial vehicles (UAVs) are robots which can fly remotely or autonomously without a human operator. Thanks to recent technological advances, UAVs are developed for wide rage of applications. Due to their relatively small size, ease of use, and flying capability, UAVs bring continuing growth in applications from government authorities to commercial related tasks used by civilians such as border security, law enforcement, wildfire surveillance, agriculture, construction, insurance, and general cinematography [1]. A recent study shows that the global commercial UAV market size was valued at $5.8 billion in 2018, is expected to grow from $16 billion in 2020 to $30 billion by 2025 [2]. However, their characteristics such as the versatility, ease of use, cheap price as well as wide availability also bring serious security threads by malicious use for criminal activities. A recent report shows that UAVs have been used for evil purposes, such as collision hazards, deployment of explosive weapons, smuggling of illegal substances and privacy violations. To deal with these existing as well as future threads, the government have to possess the right equipment against illegal UAVs. Therefore, development of an anti-UAV system is extremely urgent.

UAV detection is the most important function in the anti-UAV system. Usually, UAVs are possible to be detected by analysing the signature of appearance captured by individual or integrated equipment, such as camera [3], radar [4], acoustic sensor [5] and radio frequency (RF) sensor [6]. Compared to other technologies, radar is able to provide long-range detection up to tens of kilometers subjecting to the target radar cross section (RCS) [8] in different conditions of light and weather with almost unaffected performance. Initially, the purpose of developing radar is to detect standard aircraft with relatively large RCS and high velocity, it is reported that is not suitable for detecting small RCS targets, flying low and slow trajectories [7]. Furthermore, it reported UAVs and birds sharing same key characteristics. The reliable classification between the two targets is very important to take countermeasures against drones is a key challenge to consider. Therefore, new radar architectures have been specifically designed for this application. Radar-based UAV classification is divided into two categories: based on handcrafted features and based on deep-learning features. The former selects feature through oriented signal processing techniques. Ren and Jiang distinguished between UAV and non-UAV targets by producing a complex log spectrum from the magnitude and phase information of the Fourier transform [9]. Torvik proposed a method to distinguish between large birds and UAVs of comparable size by using up to 12 polarimetric features and obtained an accuracy around 99% [10]. Zhang computed short time Fourier transform (STFT) from a dual band radars and then used a support vector machine (SVM) for classifying three different types of UAV [11]. The later utilizes the features generated from deep neural networks. Mendis used a deep belief network (DBN) to train spectral correlation function (SFC) generated from data captured by a S-band CW radar. This method achieved an accuracy around 97% for three different types of micro-drones. By simulation, Choi described UAVs as primitive shapes, such as cylinders, ellipsoids, as well as spheres and analyzed characteristic of whose RCSs in order to collect of 500 spectrograms [12]. A CNN model was designed to train these spectrograms to classify several types of drones obtained accuracy 93%.

In this paper, we propose the inception-residual neural network (IRNN) for target classification based on micro-Doppler radar image data. The proposed IRNN approach is analyzed by tuning the hyper-parameters to seek a trade-off point between computational complexity and accuracy. The experimental results on the Real Doppler RAD-DAR (RDRD) database indicate that our approach can detect the UAV with

978-1-7281-8065-6/20/$31.00 ©2020 IEEE

the confidence up to 99.5% .

The rest of this paper is organized as follows. In Section II, we describe in detail the Real Doppler RAD-DAR database (RDRD) that is used for performance evaluation. Our design of IRNN model for UAV detection based on micro-Doppler radar is introduced in Section III . The experimental results are given in Section IV with different scenarios and comparison to the-state-of-the-art methods. Finally, Section V concludes this paper.

II. Dataset Description

The performance of the proposed neural network-based UAV detection is evaluated through a public dataset, namely RDRD dataset, that is available in [13], [14]. Accordingly, the dataset was gathered by a practical measurement of a digital array radar system developed by the Microwave and Radar Group. The radar system works at a center carrier frequency of 8.75 GHz, and a maximum bandwidth of 500 MHz. Through a digital signal processing step, a Doppler-range image of size 4092×512 for each capture is obtained. The pixel value of the Doppler-range image is assigned in dBm. Once a target is detected based on a constant false alarm rate (CFAR) technique [15], a frame of size 11×61 around the target is captured. As a result, the dataset that consists of $17,485$ frames of size 11×61 of cars, drone, and people. The distribution of cars, drone and people in the dataset is summarized in Fig. 1.

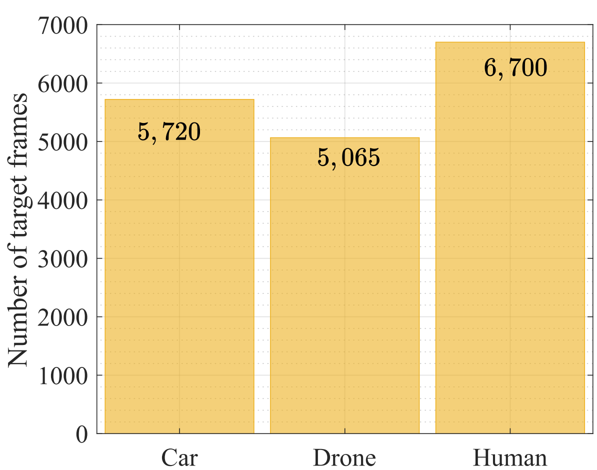

Fig. 1: Distribution of cars, drones, and humans in RDRD dataset.

It can be seen that the numbers of frames that corresponds to each target category are enough balanced for analyzing and evaluating the performance of the designed network. For illustration, Fig. 2 shows how the target-like car, drone, and human appears in the Doppler-range image frame of the radar system. In this framework, we intend to divide randomly the dataset into 80% for training and 20% for testing.

III. Proposed CNN Based UAV Detection Method

Inspired by deep learning for image processing, computer vision [16]–[18], and biomedical informatics [19], [20], CNNs has been recently exploited for handling several challenging tasks in communications [21]–[24]. Thus, in this work, we design a CNN with inception-residual architecture, as shown in Fig. 3. As can be seen in Fig. 3a, the proposed IRNN consists of three parts, including input, inception-residual, and output blocks. Accordingly, the input block contains four consecutive layers, including input, normalization, convolution, and activation layers. The input of network has to be designated with a size of 11×61, which must equal the size of target frame. The input layer is followed by a batch normalization layer, which accelerates the learning process of neural network and overcomes the vanishing gradient problem. Assume that x is input data of the bath normalization layer with a mini-batch size of B, then the normalized output is defined by the following formula:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}, \tag{1}$$

where μ_B and σ_B^2 are mini-batch mean and variance, respectively. They are defined in turn as follows:

$$\mu_B = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{2}$$

and

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2 \tag{3}$$

Despite accelerating deep network training, the batch normalization layer results in an increment of time-consuming of the prediction process in real systems.

Following the batch normalization layer is a convolution layer with K kernels of 1×1 size, which creates K channels for subsequent blocks. The two-dimensional (2D) convolution can be defined by the following equation:

$$y(i,j) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(m,n) \cdot x(i-m, j-n) \tag{4}$$

where x presents an input matrix that is convoluted with the kernel matrix h to provide an output matrix y.

In deep learning techniques, activation function is a crucial component, that determines whether the input values should be activated or not. For example, the well-known activation function, Rectified Linear Unit (ReLU), that has formula as $y = max(0, x)$ activates when x is positive and makes null for negative x. The ReLU function is computationally efficient because of quickly convergence. However, the network using ReLU function cannot perform back-propagation and cannot learn if the gradient of the function becomes zero. To overcome the dead problem of ReLU, an Exponential Linear Unit (ELU) was proposed in [25] that can still activate for negative input values. Based on that characteristic, the ELU function

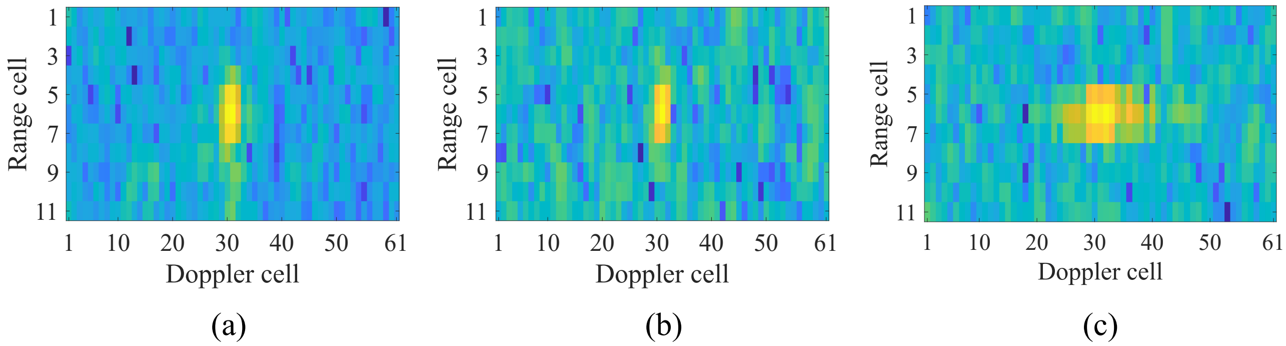

Fig. 2: Illustration of gathered frames of (a) *car*, (b) *drone*, and (c) *human* from the micro-Doppler-radar system.

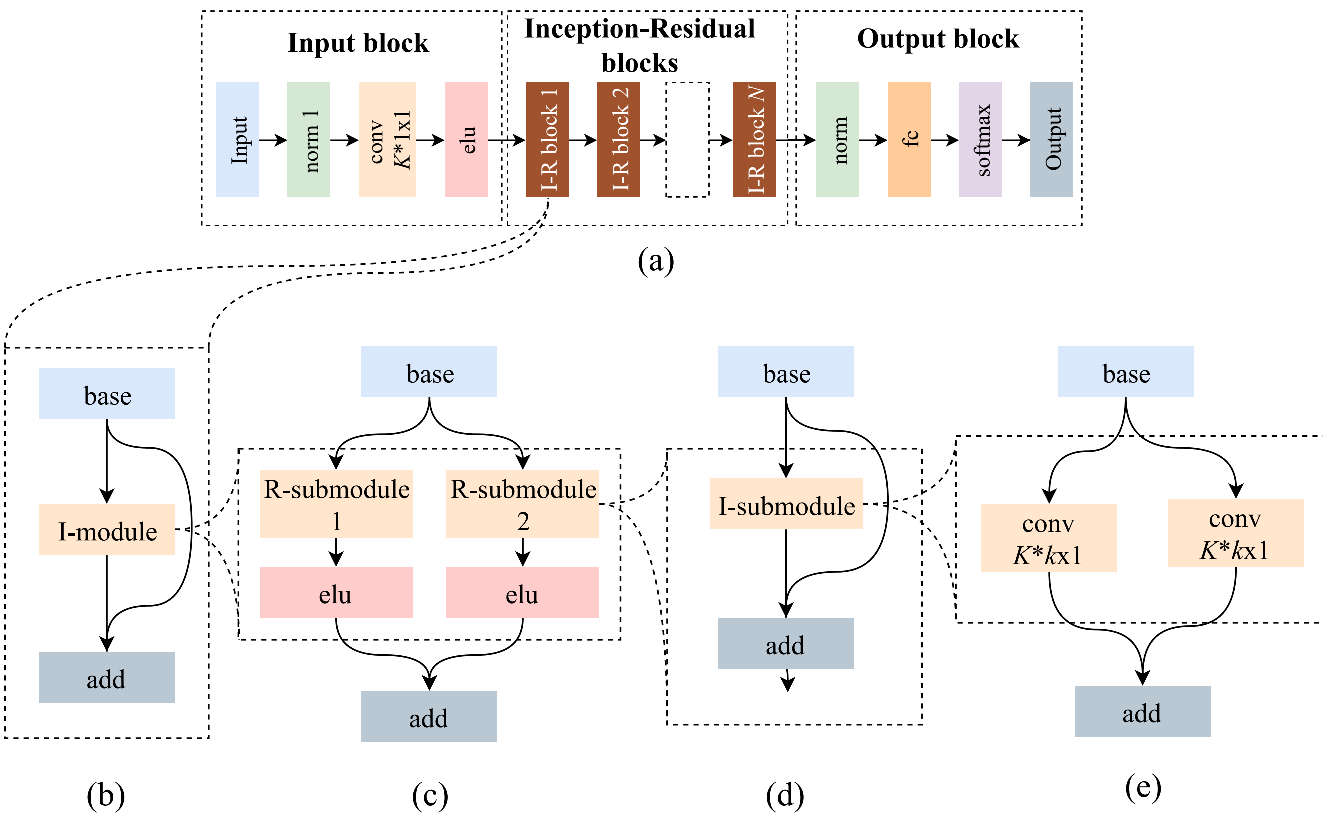

Fig. 3: Overall architecture of neural network model for micro-Doppler radar-based UAV detection.

can speed up learning rate and result in a higher classification accuracy. The ELU function can be expressed as follows:

$$y = \begin{cases} x & \text{if} x \geq 0 \\ \alpha\left(e^{x}-1\right) & \text{if} x < 0 \end{cases} \tag{5}$$

where α is a scale value between 0 and 1. The derivative of the ELU function is:

$$y' = \begin{cases} 1 & \text{if} x \geq 0 \\ y+\alpha & \text{if} x < 0 \end{cases} \tag{6}$$

The derivative of ELU is simple enough that accelerates the training process.

The inception-residual blocks (I-R blocks) are consecutively connected to each other from input to output blocks. Each block, as shown in Fig. 3b, has a structure of an inception module (I-module) combined with skip-connection that prevents the vanishing problem of network. There are two branches inside of the I-module, each of which consists of residual sub-module (R-submodule) followed by an ELU activation function layer, as shown in Fig. 3c. R-submodules 1 and 2 have the same number of filters, but they are assigned with different filter sizes. For example, in our study, $k = 5$ is for R-submodule 1 and $k = 7$ for the R-submodule 2. Subsequently, each R-submodule contains an inception submodule (I-submodule) and its skip-connection (Fig. 3d). Then, each I-submodule is constructed by two branches of convolution layers. If the first branch is a convolution layer with K filters of size $k \times 1$, then the second one has one of size $1 \times k$, as shown in Fig. 3e. This design facilitates the reduction of network trainable parameters. The proposed architecture has several advantages as follows:

- reduction of trainable parameters,
- prevention of vanishing and over-fitting problem by using the residual connection,
- learning more robustness feature by using inception module and ELU function,
- increment of training rate thanks to batch normalization.

IV. Experiment Results and Discussion

In this work, our network model is evaluated in terms of classification accuracy, time-consuming, and number of trainable parameters with different variants of filter number, and I-R block number. Each variant of the proposed model is trained with the dataset mentioned above. The training option is set with initial training rate of 0.01, which will be dropped 50% for each two epochs. The training process is executed in 20 epochs by a computer with the hardware configuration of Intel(R) Core(TM) i5-8500 CPU@3.0GHz, RAM 16GB, and GPU RTX 2070-Super 8GB.

A. UAV Detection Performance with Different Numbers of Filters

The CNN network model is designated with two I-R blocks, and different numbers of filters, such as $\{8, 16, 32, 64\}$. Many research works have confirmed that the more filters are designed, the higher the classification accuracy of the neural network is. However, due to a small number of output classes (3 classes), our model yields the highest classification accuracy when it is designated by 16 filters in convolution layers, as presented in bold in Table I.

TABLE I: UAV Detection Performance with Different Numbers of Filter.

No. filters	**Classification Accuracy (%)**			
	Average	**Car**	**Drone**	**Human**
8	97.8	95.6	98.0	99.7
16	**98.5**	**97.0**	**98.5**	**99.7**
32	98.34	96.7	98.4	99.7
64	98.2	96.5	98.2	99.7

B. Comparison Results

Competition between our proposed model with other existing ones is performed. In same task of target classification, we consider three models, namely NasNet-Mobile [26], MobileNetV2 [27], and DopplerNet [14], which have been evaluated on the same micro-Doppler radar image dataset.

The NasNet-Mobile model, whose input has a size of 224×224, is trained on more than a million images from the ImageNet database [28]. It can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. In this comparison, we modify the input size of NasNet-Mobile to be of 11×61 and the output size to be of 3 in order to satisfy the range-Doppler image dataset.

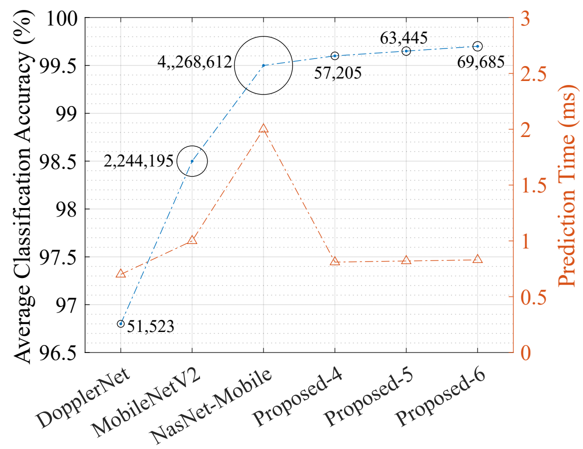

Fig. 4: Accuracy, time-consuming, and structural comparison between proposed models with other existing models.

The MobileNetV2 is a lightweight model that can improve the performance of mobile models on multiple tasks. The MobileNetV2 architecture is composed of residual structures, where the input and output of the residual blocks are bottleneck layers. It is similar as for the NasNet-Mobile model, the input and output sizes of the MobileNetV2 are modified to be of 11×61 and 3, respectively.

DopplerNet is a simple network, which consists of one convolution layer with 32 filters, filter size of 3×3, and four fully connected layers, in which three first ones have an output size of 64 and last one has an output size of 3 that corresponds to the number of target class names. The DopplerNet model has been evaluated on the same dataset that we perform in this paper. However, in that work, the network performance is only analyzed according to concatenated frames. In this comparison, we select three variants of the proposed model for competing with other models. These variants include the models of 4 I-R blocks (namely Proposed-4), 5 I-R blocks (namely Proposed-5), and 6 I-R blocks (namely Proposed-6).

The numerical results shown in Fig. 4 indicate that the DopplerNet model has the smallest number of learnable parameter resulting in the fastest prediction time (0.7 ms) but obtaining the lowest average classification accuracy (96.8%). The MobileNetv2 model that has larger structure than DopplerNet achieves higher accuracy (98.5). However, it rises more computational complexity that leads to lower prediction time (1 ms) than DopplerNet and proposed ones. NasNet-Mobile has approximately 4.3 million parameters that are two times greater than MobileNetV2; therefore, it can more accurately classify the car, drone, and human but takes a lower prediction process (up to 2ms). In contrast, with the clever design of inception-residual combination, despite having small learnable parameters (approximately 55k parameters, that are comparable with the number of parameters of DopplerNet, approximately 52k parameters), our proposed models yield higher classification accuracy than three above-mentioned ones. Additionally, the proposed models execute classification in a short time (approximately 0.8 ms). Moreover, it can be seen in Fig. 4 that the proposed network that is designed by more I-R blocks achieves higher accuracy than that of fewer ones. However, more I-R blocks result in a larger network and longer prediction time.

V. Conclusion

In this study, we have proposed a novel neural network model, which has been designed by using a combination of inception and residual conceptions. The proposed network performance in terms of classification accuracy, time-consuming, and computational complexity has been demonstrated via experiment on the RDRD dataset. In addition, various variants of our model are taken into account to compare with other existing models. Accordingly, the numerical results show that the model with 16 convolution filters yields the highest accuracy. Regarding the impact of the number of I-R blocks on the classification performance of the network, we can confirm that the more the I-R blocks are designed, the more accurately the model classifies. In comparison, despite having a small number of learnable parameters, our chosen networks achieve higher classification accuracy and faster prediction time than other considered models. For future works, we intend to extend the model for multimodal (video, radar, audio, and RF data) networks, which can enhance the drone detection and classification performance of surveillance systems.

Acknowledgment

The ASEAN IVO project [29] "An energy efficient, self-sustainable, and long range IoT system for drought monitoring and early warning" was involved in the production of the contents of this publication and financially supported by NICT.

References

[1] Research, G.V. Commercial UAV Market Analysis By Product (FixedWing, Rotary Blade, Nano, Hybrid), By Application (Agriculture, Energy, Government, Media and Entertainment) In addition, Segment Forecasts to 2022. 2016. Available online: https://www.grandviewresearch.com/industry-analysis/commercial-uavmarket (accessed on 24 April 2019).

[2] https://www.grandviewresearch.com/industry-analysis/global-commercial-drones-market

[3] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma, and M. Blumenstein, "Drone Detection in Long-Range Surveillance Videos," in *Proc. 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Taipei, Taiwan, 2019, pp. 1-6.

[4] J. S. Patel, F. Fioranelli, and D. Anderson, "Review of radar classification and RCS characterisation techniques for small UAVs or drones," *IET Radar, Sonar & Navigation*, vol. 12, no. 9, pp. 911-919, Sept. 2018.

[5] S. Al-Emadi, A. Al-Ali, A. Mohammad, and A. Al-Ali, "Audio Based Drone Detection and Identification using Deep Learning," in *Proc. 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Tangier, Morocco, 2019, pp. 459-464.

[6] M. F. Al-SaD, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF-based drone detection and identification using deep learning approaches: An initiative towards a large open source drone database," *Future Generation Computer Systems*, vol. 100, pp. 86–97, 2019.

[7] P. Molchanov, R. Harmanny, J. De Wit, K. Egiazarian, and J. Astola, "Classification of small UAVs and birds by micro-Doppler signatures", *International Journal of Microwave and Wireless Technologies*, 2014, 6, 435–444.

[8] E. F. Knott, J. F. Schaeffer, and M. T. Tulley, *Radar Cross Section*. SciTech Publishing: New York, NY, USA, 2004.

[9] J. Ren and X. Jiang, "Regularized 2-D complex-log spectral analysis and subspace reliability analysis of micro-Doppler signature for UAV detection," *Pattern Recognition*, 2017, 69, pp. 225–237.

[10] B. Torvik, K. E. Olsen, and H. Griffiths, "Classification of birds and UAVs based on radar polarimetry," *IEEE Geoscience and Remote Sensing Letters* , vol. 13, no. 9, pp. 1305-1309, Sept. 2016.

[11] P. Zhang, L. Yang, G. Chen and G. Li, "Classification of drones based on micro-Doppler signatures with dual-band radar sensors," in*Proc. 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL)*, Singapore, Nov. 2017, pp. 638–643.

[12] B. Choi and D. Oh, "Classification of drone type using deep convolutional neural networks based on micro-Doppler simulation," in *Proc. 2018 International Symposium on Antennas and Propagation (ISAP)*, Busan, Korea (South), 2018, pp. 1-2.

[13] "RDRD database - Kaggle". Available at https://www.kaggle.com/iroldan/real-doppler-raddar-database, accessed June 2020

[14] I. Roldan et al., "DopplerNet: a convolutional neural network for recognising targets in real scenarios using a persistent range–Doppler radar," *IET Radar, Sonar & Navigation,* vol. 14, no. 4, pp. 593–600, Apr. 2020,

[15] Ed. M. A. Richards, J. A. Scheer and W. A. Holm, *Principles of Modern Radar: Basic principles.* Institution of Engineering and Technology, 2010.

[16] T. Huynh-The, C. Hua and D. Kim, "Encoding Pose Features to Images With Data Augmentation for 3-D Action Recognition," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3100-3111, May 2020

[17] T. Huynh-The, C. Hua, N. Tu, D. Kim , "Learning 3D spatiotemporal gait feature by convolutional network for person identification," *Neurocomputing*, vol. 397, pp. 192-202, Jul. 2020.

[18] T. Huynh-The, C. Hua, T. Ngo, D. Kim, "Image representation of pose-transition feature for 3D skeleton-based action recognition," *Information Sciences*, vol. 513, pp. 112-126, Mar. 2020.

[19] C. Hua, T. Huynh-The and S. Lee, "Convolutional Networks with Bracket-Style Decoder for Semantic Scene Segmentation," in *Proc. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, 2018, pp. 2980-2985.

[20] C. Hua et al., "Bimodal learning via trilogy of skip-connection deep networks for diabetic retinopathy risk progression identification," *International Journal of Medical Informatics*, vol. 132, pp. 103926, Dec. 2019.

[21] T. Huynh-The, C. Hua, J. Kim, S. Kim and D. Kim, "Exploiting a low-cost CNN with skip connection for robust automatic modulation classification," in *Proc. 2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1-6.

[22] T. Huynh-The, C. Hua, Q. Pham and D. Kim, "MCNet: An Efficient CNN Architecture for Robust Automatic Modulation Classification," *IEEE Communications Letters*, vol. 24, no. 4, pp. 811-815, April 2020.

[23] V.-S. Doan, T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Learning Constellation Map with Deep CNN for Accurate Modulation Recognition," *arXiv preprint arXiv: 2009.02026*, 2020.

[24] T. Huynh-The , V.-S. Doan, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Chain-Net: Learning Deep Model for Modulation Classification Under Synthetic Channel Impairment," *arXiv preprint arXiv: 2009.02023*, 2020.

[25] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," Nov. 2015, arXiv:1511.07289. [Online]. Available: https://arxiv.org/abs/1511.07289

[26] B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 8697-8710.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 4510-4520,

[28] ImageNet. http://www.image-net.org

[29] ASEAN IVO. http://www.nict.go.jp/en/asean$_ivo/index.html$