# An Evaluation of CNN using Deep Residual Learning for Modulation, 5G, LTE, and WLAN System Classification

Teruji Ide†
*Dept. of Electrical and Electronic Engineering*
*National Institute of Technology, Kagoshima College*
Kirishima-shi, Kagoshima, Japan
t-ide@kagoshima-ct.ac.jp

Tadatomo Sato†
*Dept. of Electrical and Electronic Engineering*
*National Institute of Technology, Kagoshima College*
Kirishima-shi, Kagoshima, Japan
sato@kagoshima-ct.ac.jp

Rozeha A. Rashid††
*Faculty of Electrical Engineering*
*Universiti Teknologi Malaysia*
Johor Bahru, Johor, Malaysia
rozeha@utm.my

Leon Chin††
*Faculty of Electrical Engineering*
*Universiti Teknologi Malaysia*
Johor Bahru, Johor, Malaysia
leonchinmpm@gmail.com

M.A. Sarijari††
*Faculty of Electrical Engineering*
*Universiti Teknologi Malaysia*
Johor Bahru, Johor, Malaysia
madib@utm.my

Rubita Sudirman††
*Faculty of Electrical Engineering*
*Universiti Teknologi Malaysia*
Johor Bahru, Johor, Malaysia
rubita@utm.my

***Abstract*—In this study, we investigate and present a deep residual learning for modulation and system classification. The simulation results show the degradation problem that was exposed due to an increase in network depth and the saturation of accuracy in the conventional CNN; however, the proposed CNN has no such degradation. Therefore, the processing burden of the conventional CNN is much larger than the proposed CNN. In the simulation results, the proposed CNN framework achieves better system (5G, LTE, and WLAN) classification accuracy as the conventional CNN framework when reducing the processing burden in the proposed one. The better simulation results are shown by adjustment of the parameters using the proposed method in the case of 5G, LTE, and WLAN systems.**

***Keywords—CNN, cognitive radio, deep residual learning, modulation & system classification***

## I. Introduction

Recently, advanced development of mobile communication systems has been increasing in cognitive radio (CR) techniques [1] based on software defined radio (SDR) technology. To perform effectively utilizing frequencies, CR users should recognize the surrounding radio environment, take the required measurements, and make adequate decisions in order to use frequency resources. On the other hand, spectrum sensing must be able to identify and distinguish between the primary user (PU) and other secondary users (SUs) using only energy detection; however, it is not easy to distinguish difference among the systems. For example, modulation classification (MC) [2][3][4][5] is a fascinating method to solve this problem in terms of accuracy of sensing.

Moreover, the Fifth-Generation Mobile communications system (5G) has become the focus of the new generation of communication all over world as an extensive advancement of the existing mobile communication systems based on intensive requirements of market trends to the mobile communication systems in recent years. However, because of the shortage of frequencies for 5G, 6G, and other conventional systems, developments for frequency sharing with a wireless resource secured dynamically have actively been improving. To avoid interfering from SUs to PU, SUs have to recognize surrounding radio environments.

Because of this, sensing and radio environment map (REM) are useful as shown in Fig.1. Creating REM needs electric field strengths measurement, received power detection, and presence of PU's because it is easy to detect. However, it needs the information what wireless systems and modulation types are PU's. Therefore, their modulation and system (5G, LTE (Long Term Evolution), and WLAN (Wireless Local Area Network), etc.) classification is needed.

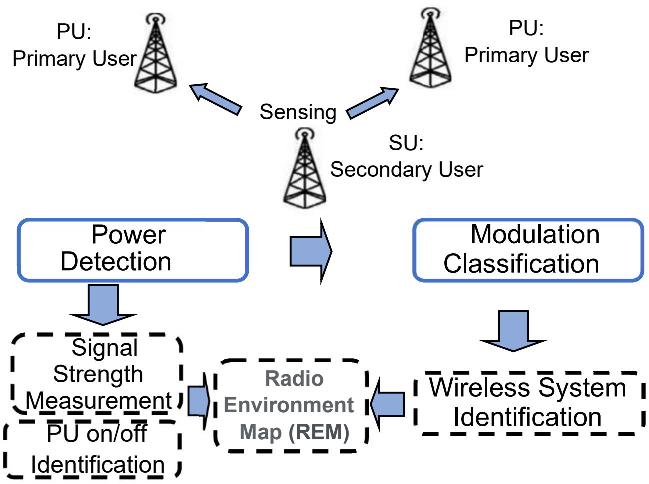


Fig. 1. Modulation & systems classification and Radio Environment map (REM)

Machine learning [2] is a method of implementing the identification of different types of MC and mobile communication systems by learning from training data. Deep Neural Networks (DNN) have played a significant role in radio communication area in the past, as well as in video, speech, and image (picture) processing. Moreover, the Convolutional Neural Network (CNN) [5] [6] has been one of the useful methods in deep learning for image processing as a powerful tool; however, it is not easy to stack more layers in spite of using the CNN. Usually, deeper networks are required to have the ability to be convergent;

however, a degradation problem may occur as the network depth increases and with a saturation of accuracy of sensing. The activation functions are differentiated in every layer to calculate the gradient. First, the differential operations are useful because the differences of accuracy between inputs and outputs (one layer) in the shallow layers near the input layer are large. The characteristics of propagation of layer to layer are multiplying, so the decreasing characteristics are exponential; that is to say, the gradient disappears by increasing of the layers. To solve the above problem, we introduce a deep residual learning method [6]. It features underlying mapping with a nonlinear H(x) function from input to output. Instead of H(x) using a nonlinear function F(x), which is defined as H(x)-x, the method arithmetically adds x to F(x) at the output of the second weight layer, as shown in Fig. 2. The novelty of the method in the past. We compared the characteristics of a normal CNN and our proposed CNN using a deep residual learning (ResNet) method. The performance evaluation revealed a significant accuracy advantage and a reduction in processing complexity.

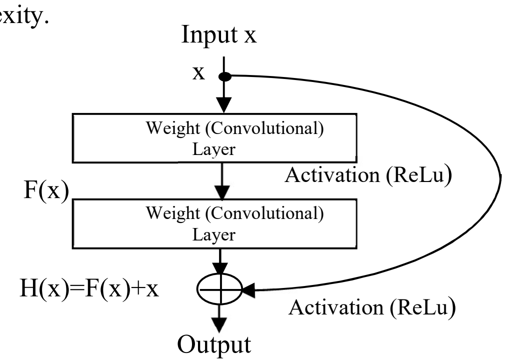


Fig. 2. A CNN using a deep residual learning block diagram.

## II. ARCHITECTURE OF CNNS USING DEEP RESIDUAL LEARNING

Fig. 3 shows a conventional (normal) CNN block diagram for the purpose of comparing the proposed CNN using deep residual learning framework. In Fig. 3, the seven combinations (in turn and serially cascaded) of the convolutional layer, the batch-normalization layer, the activation layers whose function is ReLu (if the magnitude of the input signal is larger than zero, the output signal is the same as the input one; otherwise, the input signals are replaced with zero), and the pooling layer that extracts the maximum value in the feature maps are serially cascaded. The filter size is 8 and the numbers of filters are 16 (first stage), 24, 32, 48, 64, 96, and 256 (last stage) in turn. Fig. 6 shows the proposed CNN using deep residual learning. In Fig. 6, the combination (in turn and serially cascaded) of the convolutional layer, the batch-normalization layer, and the activation layers whose function is ReLu; however, there are only these three combinations and each 16 filters to reduce the processing burden.

In addition, the proposed CNN is a combination of a normal CNN and a ResNet (CNN using deep residual learning) as shown in Fig. 2. A normal (conventional) CNN in Fig. 3 has the advantage of faster convergence of learning compared to that of a ResNet; however, a normal CNN has saturation of accuracy, but a ResNet has no saturation of accuracy. Considering these drawbacks and merits, the proposed CNN is the combination of both types of CNN and the block diagram is shown in Fig. 6. In Fig. 6, we add some CNN combinations in Fig. 2 (Res Net), as mentioned above, to the combinations of normal CNN (convolutional layer, batch-normalization layer, activation layer whose function is ReLu, and pool layer) whose filter size equals eight, and the numbers of filters are 16 (first stage) to 256 (last stage), respectively. On the whole, the combination of the conventional CNN and ResNet using CNN has the advantage of the number of layers (less than the architecture to take the same characteristics) and the faster convergence time to take the required validation accuracy.

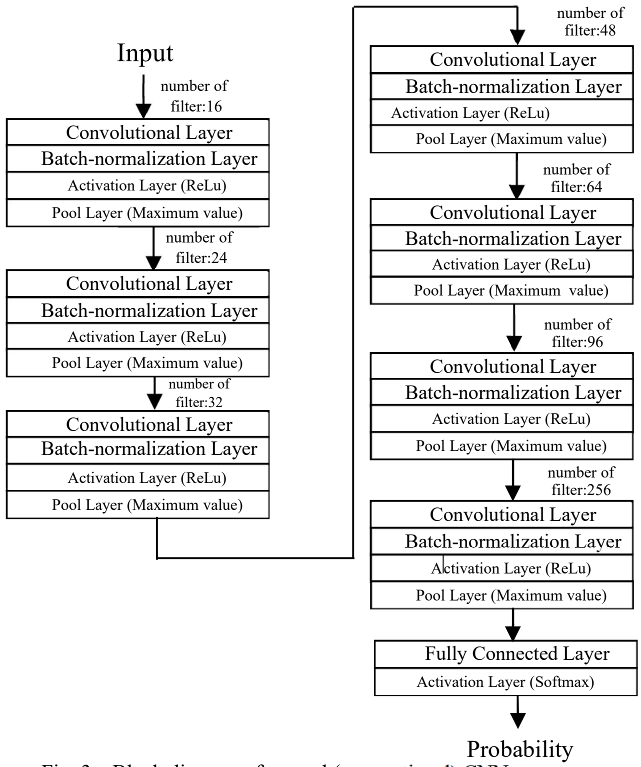


Fig. 3. Block diagram of normal (conventional) CNN

## III. SIMULATION RESULTS AND DISCUSSIONS

### A. Modulation classifiction

Table 1 (A) shows the simulation parameters (the modulation types are 64QAM, 16QAM, 8PSK, QPSK and BPSK, AWGN channel (S/N=20dB)). Table 1 (B) contains the results to enable a comparison of the use of the conventional CNN and the proposed CNN using deep residual learning.

Fig. 4 shows the block diagram of CNN using only Res Net (11 stages). The simulation results using this block diagram show in Fig. 5. In Fig. 5, the training accuracy (CNN using only Res Net) is approximately 20 % (not shown in) and the validation accuracy is less than 20%. The more ResNet is being used, the more characteristics of the validation accuracy is higher than when using less ResNet characteristics. It is possible to achieve 90% validation accuracy using only ResNet. However, using only ResNet has lower convergence speed than that of in the case of not using ResNet. To achieve high validation accuracy, the machine learning unit need a lot of processing burden in the


†The authors are with National Institute of Technology, Kagoshima College, 1460-1 Shinkou Hayato-cho, Kirishima-shi, Kagoshima 899-5193, Japan.

††The authors are with Universiti Teknologi Malaysia, 1310 Johor Bahru, Johor, Malaysia.

case of using ResNet. To achieve better performance such as more than 70% of training and validation accuracy, it is supposed that the number of ResNet is larger than that of ResNet stages (for example: ten times) in Fig. 4.

The simulation results using the conventional CNN shown in Fig. 3 show in Fig, 5. A better performance of the training accuracy (almost 100%, not shown) can be achievable by increasing of CNN; however, the validation accuracy remains approximately 60% due to over fitting. On the other hand, a normal (conventional) CNN has the advantage of faster convergence of learning as shown in Fig. 5. This means that an over-fitting in this machine learning occurs. Considering these drawbacks and merits, the proposed CNN is the combination of both types of CNN and the block diagram is shown in Fig. 6. The simulation results of characteristics of the block diagram in Fig.6 are show in Fig. 5. The simulation results in Fig5 show the better performance of a faster convergence and approximately 70 to 80% of a validation accuracy.

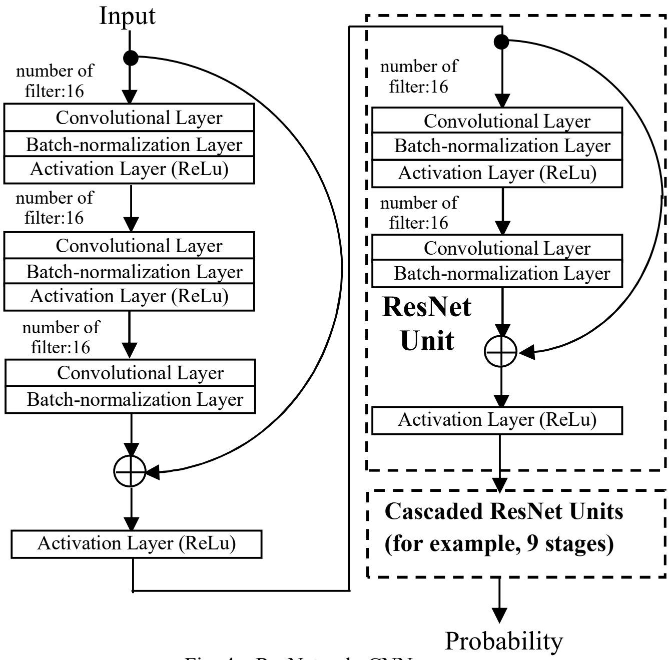


Fig. 4. ResNet only CNN

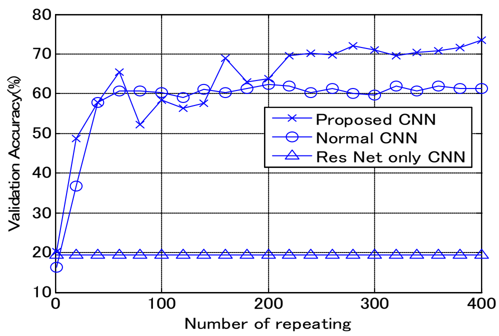


Fig. 5. Validation accuracy

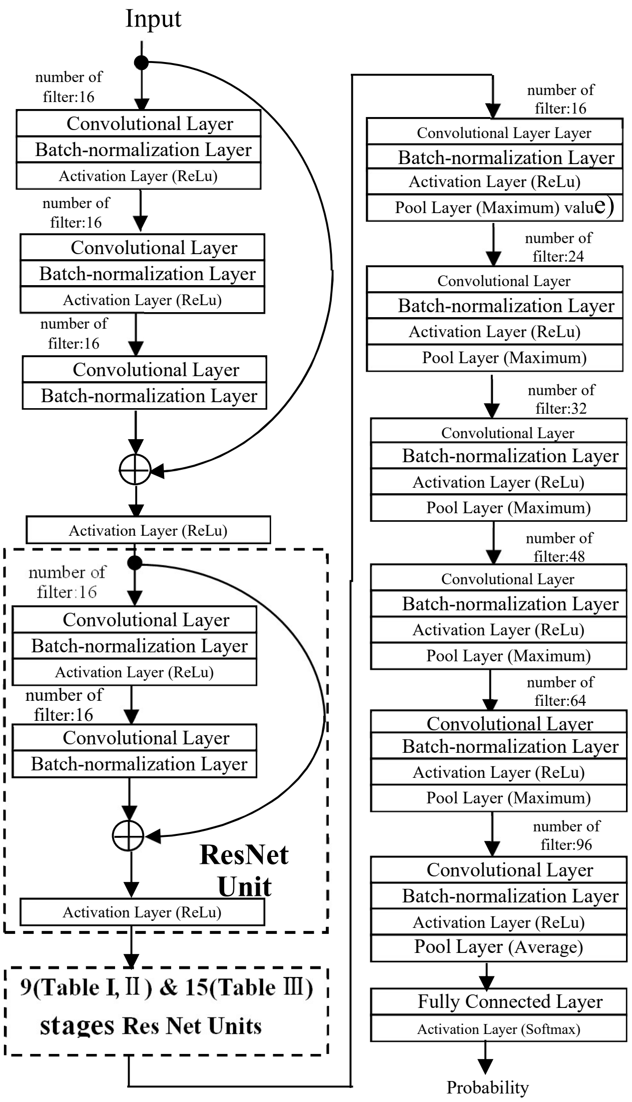


Fig. 6. Block diagram of the proposed method (Table III: adding one layer (number of filter 64) between the fifth and sixth layer and changing the number of filters of 96 to 128 for the last layer).

TABLE I. SIMULATION PARAMETERS AND RESULTS

(A) PARAMETERS

| | |
|---|---|
| **Modulation Type** | 64QAM, 16QAM, 8PSK<br>QPSK, BPSK |
| **Channel** | AWGN (S/N=20dB) |
| **Data** | Training:80%, Validation:10%, Test:10% |

(B) RESULTS

| **Modulation Type** | **Accuracy (Probability)** | | |
|---|---|---|---|
| | ***Proposed CNN*** | ***Conventional CNN*** | ***ResNet only CNN*** |
| 64QAM | 63.6% | 61.5% | 0% |
| 16QAM | 64.3% | 54.4% | 0% |
| 8PSK | 79.1% | 48.4% | 0% |
| QPSK | 69.2% | 45.6% | 0% |
| BPSK | 96.2% | 100% | 100% |

The simulation results of proposed method in Fig.7 (the block diagram shows in Fig.6) is different from the above (conventional CNN). In Fig.7, the converged final training and validation values are almost the same; that is to say, the value is approximately 70 to 80 %. It is supposed that the difference in Fig. 5 is caused by the problem of “Vanishing Gradient.”

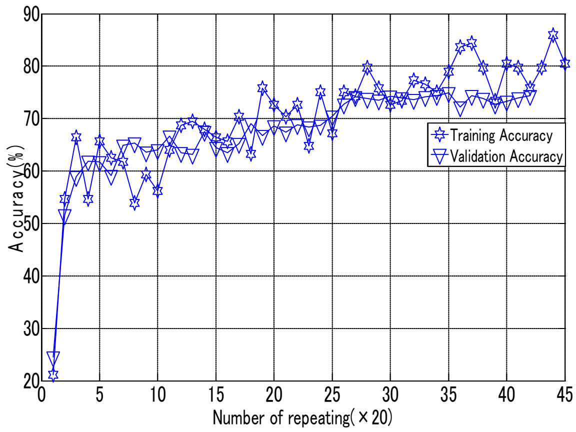


Fig. 7. Validation accuracy (proposed method)

Fig,8 shows the results of confusion matrix compared among these five modulations in the block diagram of Fig. 6 (the simulation results (validation accuracy) are shown in Fig. 7). In Fig. 8, the lowest row shows the ratio of the expected data (CNN outputs: modulation types) to the true values (test data: modulation types), and the right side column shows the ratio of the true values to the expected data.

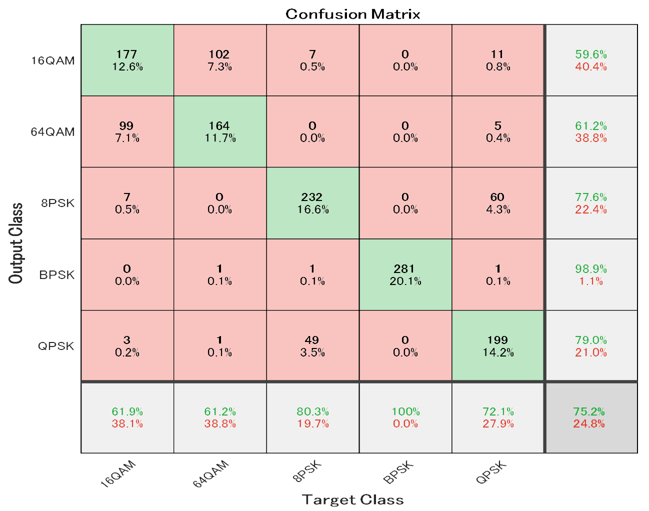


Fig. 8. Confusion matrix (proposed method)

Fig.9 shows the results of the validation (training) accuracy (convergence) characteristics for different dropout rates using the proposed method. In Fig. 9, the results show that the adequate dropout rate is 0.1 to 0.3 when using the proposed method. Because of the results, there is no influence of overfitting using the proposed method. In this simulation, the dropout operation is applied to not only "Fully Connected Layer" but also the all CNN layers.

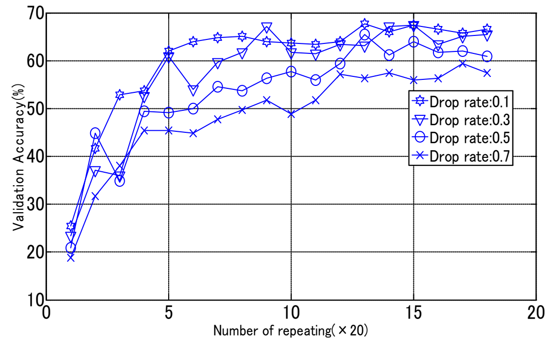


Fig. 9. Characteristics for the proposed method (different dropout parameters)

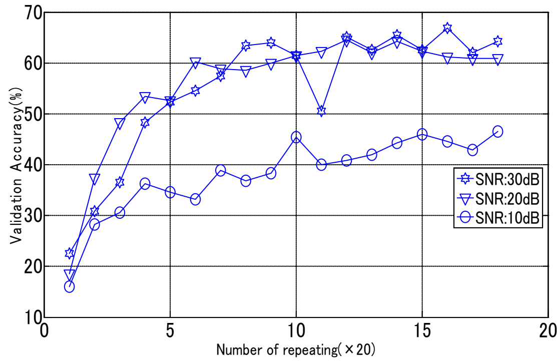


Fig. 10. Characteristics for the proposed method (different SNR parameters)

In Fig,10, analyses show that the proposed method yields a classification accuracy higher than 70% at varying SNR conditions ranging from 10dB to 30dB. Compared with the conventional method in Table Ⅰ, the proposed method effectively improves the problem of poor classification performance for the above range of SNRs and possesses better robustness. Comparing Fig. 3 and Fig. 6, the processing burden of the conventional CNN is much larger than that even though the results of the probability classification for the modulation types are almost the same. As for the results of using the conventional CNN, the probability classifications of the modulation types are “64QAM, 16QAM, 8PSK, QPSK, and BPSK” and “61.5%, 54.4%, 48.4%, 45.6%, and 100%,” respectively. On the other hand, as for the results of using the proposed CNN using deep residual learning, the probability classifications of the modulation types are “64QAM, 16QAM, 8PSK, QPSK, and BPSK” and “63.6%, 64.3%, 79.1%, 69.2%, and 96.2%,” respectively.

The Simulation results of ResNet only are not better than the others as shown in Table Ⅰ. It is supposed that ResNet requires a lot of CNN layers for a better performance. In this simulation (not used for GPU), 700 data are repeated and the training, validation, and testing data are 80%, 10%, and 10%, respectively. The total training time is approximately three hours (in the case of ResNet [5], it took 20 hours).

In the proposed method, the results using the number of filter (ResNet) of 64 (validation accuracy is approximately

85%) is better than that of the filter number 16 (validation accuracy is approximately 80%). They are adequate simulation results; however, the former case has to require a processing burden compared with the latter one. Considering the simulation results, a degradation problem was exposed with the increased network depth, wherein accuracy becomes saturated in the conventional CNN; however, the proposed CNN suffers no such degradation. As for the proposed CNN and the conventional CNN, judging from the results of each modulation classification, a small portion of the 8PSK data are misclassified as QPSK, and 16QAM and 64QAM tend to be misclassified as each other. In this simulation in the proposed CNN, 16 filters are used of size eight only; however, validation is needed as to whether or not there will be a saturation of accuracy when varying the number of filters and their size. In addition, validation will be needed for an increase in the probability (accuracy) when increasing the ResNet and normal CNN blocks in the proposed CNN. The proposed CNN method achieved an overall validated of classification accuracy of 70 to 80% at a range of 20 to 30 dB SNR and the dropout rate of 0.0 (no dropout) or 0.1 over a class of 5 digital modulation schemes compared with the conventional method which achieves an accuracy of approximately 60%.

### *B. System (5G, LTE, WLANs) classifiction*

Table II (B) shows the results of these four systems (LTE and WLAN) classification using the block diagrams of Figs. 3(conventional) and 6(proposed) in the case that these SNR are 30dB to -20dB. In the case where SNR is 0dB, all the results of system classifications when using the proposed method are 100%). On the other hand, the results of proposed method are better than the conventional one in the case of poor SNR values in the case of -20dB as shown in Table 1 (B). Compared with the conventional method, the proposed methods improve the problem of poor classification performance for the above range of SNRs and possesses better robustness. Fig.11 shows the results of the validation (training) accuracy (convergence) characteristics for different dropout rates using the proposed method. In Fig. 11, the results show that the adequate dropout rate is 0 (or no need to have the dropout operation) or 0.1 when using the proposed method (the final validation result reaches 100%, not shown in Fig. 11). Because of the results, there is no influence of overfitting using the proposed method. In this simulation, the dropout operation is applied to not only the "Fully Connected Layer" but also the all CNN layers. In these simulation results, as shown in Fig. 11, it is assumed that there is no inference of overfitting.

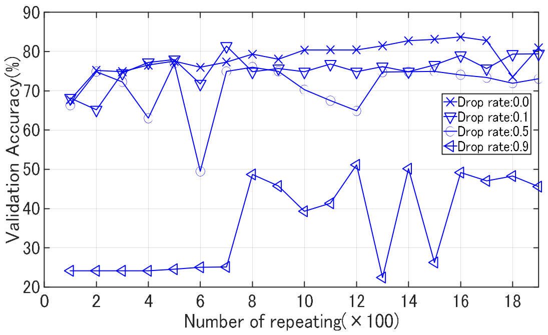


Fig. 11. Characteristics (dropout) for the proposed method (LTE&WLAN) (SNR:10dB)

Table Ⅲ(A) shows the simulation parameters (the systems are WLAN (ax) (downlink), WLAN(ax)(uplink), WLAN(ah)(downlink), LTE(downlink), and 5G (downlin-k) AWGN channel(SNR=-20 to 30dB)).

Table Ⅲ (B) shows the results of classification probability compared among these five systems in the block diagram of Fig. 6 (the parameters are shown in Table Ⅲ). In Table Ⅲ in the case of SNR is 10dB to 30dB, all the results of system classifications are 100% (and more, the same results are obtained in the case where of these SNR are higher than that). On the other hand, the results of the proposed method are better than the conventional one in the case of poor SNR values in the case of -20dB (the results are shown in Table Ⅲ (B)). Compared with the conventional method, the proposed method improves the problem of poor classification performance for the above range of SNRs and possesses better robustness. In LTE systems, the subcarrier spacing is fixed (15kHz). On the other hand, the subcarrier spacings are not fixed (15kHz times n-th power of 2, n equals 0 to 4, 15kHz to 240kHz) and the number of slots equals 1 to 16 in 5G systems. As a whole, the classification of 5G and LTE is very difficult. To distinguish between LTE and 5G, we have added one layer (number of filter 64) between the fifth and sixth layer and changed the number of filters of 96 to 128 for the last layer; in addition, we have changed the number of ResNet units of 11 to 17 in Fig. 6. In the proposed method, it is supposed that the combination of conventional CNN and ResNet units should adjust together.

TABLE II. SIMULATION PARAMETERS AND RESULTS

(A) PARAMETERS

| | |
|---|---|
| **Classification System** | LTE(E-UTRA test models(E-TM 1.1) (QPSK)<br>802.11ax(downlink, BPSK,16QAM)<br>802.11ax (uplink, BPSK,16QAM)<br>802.11ah(downlink,BPSK,16QAM) |
| **Channel** | AWGN(SNR= -20～30dB) |
| **Data** | Training:80%, Validation:10%, Test:10% |

(B) RESULTS

| **System Type** | **Accuracy (Probability)** | |
|---|---|---|
| | ***Proposed CNN*** | ***Conventional CNN*** |
| LTE(downlink) | 100%(SNR:0～30dB)<br>33.1%(SNR:-20dB) | 100%(SNR:10～30dB)<br>35.5%(SNR:-20dB) |
| WLAN(802.11ah)<br>(downlink) | 100%(SNR:0～30dB)<br>40.1%(SNR:-20dB) | 100%(SNR:10～30dB)<br>29.8%(SNR:-20dB) |
| WLAN(802.11ax)<br>(downlink) | 100%(SNR:0～30dB)<br>100%(SNR:-20dB) | 100%(SNR:10～30dB)<br>100%(SNR:-20dB) |
| WLAN(802.11ax)<br>(uplink) | 100%(SNR:0～30dB)<br>40.2%(SNR:-20dB) | 100%(SNR:10～30dB)<br>36.8%(SNR:-20dB) |

TABLE III. SIMULATION PARAMETERS AND RESULTS

(A) PARAMETERS

| | |
|---|---|
| **Classification System** | (1)LTE(E-UTRA test model(E-TM 1.1) ) (QPSK)<br>(2)5G(ETSI TS 138 141-1 V15.5.0)<br>test model 3.1a (NR-FR1-TM3.1a) (256QAM)<br>(3)802.11ax (downlink, 16QAM)<br>(4)802.11ax (uplink, 16QAM)<br>(5)802.11ah (downlink, 16QAM) |
| **Channel** | AWGN(SNR= -20～30dB) |
| **Data** | Training:80%, Validation:10%, Test:10% |

(B) RESULTS

| **System Type** | **Accuracy (Probability)** | |
|---|---|---|
| | ***Proposed CNN)*** | ***Conventional CNN*** |
| 5G(down link) | 100%(SNR:10～30dB)<br>30.1%(SNR:-20dB) | 100%(SNR:10～30dB)<br>100%(SNR:-20dＢ) |
| LTE(down link) | 100%(SNR:10～30dB)<br>16.7%(SNR:-20dＢ) | 100%(SNR:10～30dB)<br>0.0%(SNR:-20dＢ) |
| WLAN(802.11ah)<br>(down link) | 100%(SNR:10～30dB)<br>26.3%(SNR:-20dＢ) | 100%(SNR:10～30dＢ)<br>0.0%(SNR:-20dB) |
| WLAN(802.11ax)<br>(down link) | 100%(SNR:10～30dB)<br>100%(SNR:-20dB) | 100%(SNR:10～30dB)<br>0.0%(SNR:-20dB) |
| WLAN(802.11ax)<br>(up link) | 100%(SNR:10～30dB)<br>29.4%(SNR:-20dB) | 100%(SNR:10～30dB)<br>0.0%(SNR:-20dB) |

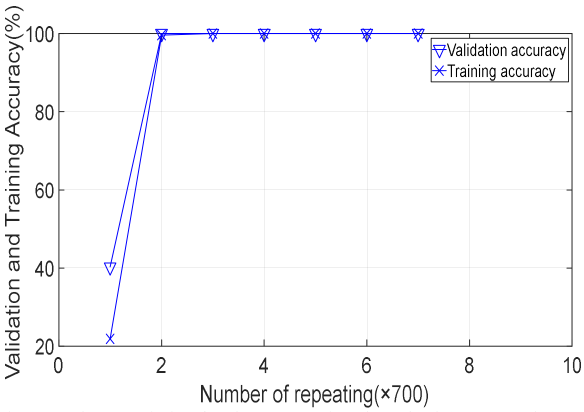

Fig. 12. Characteristics for the conventional method (SNR=10dB) (training and validation accuracy)

Fig. 12 shows the simulation results of the proposed method in Fig. 6. The training accuracy reaches approximately 100% when the value is converged when comparing among systems (5G(downlink) (256QAM), LTE (downlink)(QPSK), WLAN ah(downlink)(16QAM), WLA N ax(downlink)(16QAM), and WLAN ax(uplink)(16QA M)) using the proposed method in Table Ⅲ. In these simulation results, the training and validation accuracy reaches 100% when the value is converged in the case that SNR equals 10dB. The proposed method validates good performance in Fig. 12 (block diagram shows in Fig. 6). In Fig.12, the converged final training and validation values are almost the same; that is to say, the value is approximately 100% in the case that SNR equals 10dB. It is supposed that it is not caused by the problem of "Vanishing Gradient" because the proposed method is used. On the other hand, it causes deterioration of characteristics (training and validation accuracy) when selecting dropout in the dropout simulation (not shown in the figure of results in this paper); that is to say, the proposed method does not cause the overfitting as well as the results of the simulation of Table I and Table Ⅱ.

## CONCLUSIONS

We have presented and evaluated the proposed CNN using a deep residual learning method. As for the evaluation of the proposed method, the different dropout rates and SNR values are examined and the effectiveness of the proposed method has been ensured in this paper. As increasing the data patterns per one epoch, the validation accuracy for the proposed method can be increased. The simulation results are that the figure of validation accuracy reaches approximately 80% (modulation classification) and 100% (system classification). Using the proposed method, it is supposed that it is not caused by the problem of "Vanishing Gradient." In the case using the conventional method, the accuracy does not increase by a significant value. To ensure better performance, as the ResNet and the conventional CNN blocks are increased in the proposed CNN, an adjustment of the parameters in the CNN will be needed. In addition, a combination of long short-term memory (LSTM)[5] and the proposed CNN will be able to be used for learning long-term dependencies data.


## ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP21K14161.


## References


[1] Mitola, J. III, Cognitive Radio Architecture, John Wiley, New York, 2006.

[2] Ruslan Politanskyi ; Mykhailo Klymash, "Application of Artificial Intelligence in Cognitive Radio for Planning Distribution of Frequency Channels," 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT), 2-6 July 2019.

[3] X. Zhu and T. Fujii, "Modulation classification in cognitive radios for satellite and terrestrial systems," in IEEE International Conference on Communications Workshop, London, UK, 2015, pp. 1612–1616.

[4] X. Zhu and T. Fujii, "A modulation classification method in cognitive radios system using stacked denoising sparse autoencoder," 2017 IEEE Radio and Wireless Symposium (RWS), Phoenix, AZ, USA, 2017, DOI: 10.1109/RWS.2017.7885992.

[5] Xiaoyu Liu, Diyu Yang, and Aly El Gamal, "Deep neural network architectures for modulation classification," 2017 51st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, pp. 915-919.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, DOI: 10.1109/CVPR.2016.90.