

Article

Research on Signal Modulation Classification under Low SNR Based on ResNext Network

Binghang Zou ¹, Hanzhi Yan ², Faquan Wang ¹, Yucheng Zhou ³ and Xiaodong Zeng ^{1,*}¹ School of Electrical Engineering, Sichuan University, Chengdu 610065, China² College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an 710600, China³ Glasgow College, University of Electronic Science and Technology of China, Chengdu 610056, China

* Correspondence: zengxiaodong@scu.edu.cn

Abstract: To address the shortcomings of existing methods such as low recognition accuracy and poor anti-interference performance under low signal-to-noise ratios, this paper proposes the RFSE-ResNeXt (Residual-fusion squeeze–excitation aggregated residual for networks, RFSE-ResNeXt) network. In this paper, we improve the residual structure of the network based on the ResNeXt network and then introduce the compressed excitation structure to improve the generalization ability of the network. The improvement of the residual structure of the network leads to a good improvement in the overall recognition accuracy of the network; meanwhile, the compressed excitation structure improves the confusion phenomenon when the network faces complex signals with low signal-to-noise ratios. The experimental results show that the proposed network improves the recognition accuracy by 4% on average at a very low SNR of -10dB and reduces the misclassification of AM-DSB into CPFASK by about 27%.

Keywords: automatic modulation recognition; RFSE-ResNeXt; residual fusion; squeeze excitation



Citation: Zou, B.; Yan, H.; Wang, F.; Zhou, Y.; Zeng, X. Research on Signal Modulation Classification under Low SNR Based on ResNext Network. *Electronics* **2022**, *11*, 2662. <https://doi.org/10.3390/electronics11172662>

Academic Editor: Tadeusz A. Wysocki

Received: 28 July 2022

Accepted: 22 August 2022

Published: 25 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the field of radio communication, modulation identification techniques for communication signals have been studied and updated by many scholars, and they are of vital importance for areas such as civil and military competition [1]. Currently, there are two main types of modulation identification techniques: modulation identification based on decision theory and modulation identification based on feature engineering [2]. The former test statistics are complicated to compute and require some a priori information but simple discriminative rules; the latter feature extraction is simple and easy to compute but the discriminative rules are complicated. However, both methods currently suffer from high manual involvement, shallow extraction of signal features, and easy confusion of different signals [3]. Moreover, with the increase in signal type complexity and the development of computer technology, deep learning has been shining in this field of signal modulation recognition [4–7].

Deep learning has superior recognition accuracy compared to statistical learning and hypothesis testing methods and a better ability to learn directly from high-dimensional raw input data compared to machine learning. These advantages make deep learning a popular research direction for signal recognition techniques. In the literature [8], the authors are the first to apply the convolutional neural network model (CNN) to radio signal classification and provide a simulation dataset RML2016.10a generated based on the GNU Radio environment, on which our subsequent investigations will also be based. The literature [9], on the other hand, investigates the deep structure of modulation recognition, and an improvement in effectiveness is obtained. The advantages of airborne deep-learning-based radio signal classification over conventional algorithms are discussed in the paper [10]. Further,

the literature [7] trained different neural networks such as a baseline neural network and a residual neural network and obtained relatively high recognition rates.

The new neural network frameworks proposed afterward are pursuing reliability along with lower error rates and wider data processing [11]. Since the phenomenon of gradient disappearance and gradient explosion limits the improvement of the depth of ordinary neural networks, the phenomenon of deep network degradation persists although BatchNorm has been developed to deal with this problem [12]. It was not until the residual structure was proposed, i.e., ResNet [13], that the degradation problem was solved. However, to reduce the error rate of neural networks, the traditional methods chosen to increase the network width and depth lead to an increase in the number of hyperparameters, which increases the design difficulty and computational effort of neural networks. To solve this problem, an optimized version of the neural network ResNeXt [14] based on ResNet was proposed. ResNeXt adopts a residual structure and uses group convolution [15] instead of the three-layer convolution structure of ResNet, which can improve the accuracy of the neural network while reducing the number of hyperparameters and decreasing the parameter complexity. However, it is still not effective in the face of more complex low signal-to-noise signals.

The group convolution strategy used in the ResNeXt network itself improves the recognition accuracy but weakens the generalization ability, and the confusion phenomenon is more serious at low signal-to-noise ratios. Moreover, other existing recognition methods in radio modulation signal recognition also often have poor anti-interference performance, and the signal is easy to confuse the problem. In this regard, in this paper the RFSE-ResNeXt (Residual-fusion squeeze–excitation aggregated residual for networks, RFSE-ResNeXt) network is established. By improving the residual structure of the network, the extracted deep and shallow features are fused, and the compressed excitation structure is introduced to improve the key feature weights so that the recognition accuracy of the network can be improved when facing complex signals with low signal-to-noise ratios.

2. ResNeXt Model Structure

Xie et al. [14] proposed the ResNeXt network structure at the 2017 CVPR, which is optimized based on ResNet. The existing methods usually adopt the strategy of deepening or widening the neural network in the problem of improving the accuracy of neural network models. However, increasing the depth makes the training process more difficult and increases the difficulty of convergence of neural network parameters, while increasing the width means higher complexity and more computation, and the surge in the number of parameters often results in overfitting. To address this problem, the residual module is proposed for ResNet networks, which can effectively avoid the problem of network gradient disappearance and gradient explosion brought by the deepening of CNN networks.

The new dimensional bases are proposed so that ResNeXt performs better with the same complexity of the neural network model. The original ResNet has a base number of 2, while ResNeXt increases the base number to 32 by using parallel topology. This makes ResNeXt more accurate compared to ResNet with a close number of parameters and a significant improvement in computational efficiency.

Looking at both networks internally, the increase of the ResNeXt dimension implies an improvement of the fully connected layer, which for ResNet results in a summation of the weights \times outputs. ResNeXt, on the other hand, is slightly different because it has more parallel topologies internally and thus can be represented by the following Equation (1)

$$\text{out}(a) = \sum_{i=1}^n (\partial * C(a)) \quad (1)$$

where ∂ is the weight of different topologies, $C(a)$ is the output value of a flat identical topology, and n is the number of identical branches that a module has, i.e., the base of the model. The use of blocks can show this structure more intuitively. Figure 1 shows the block structure of ResNet, and Figure 2 shows the block structure of ResNeXt:

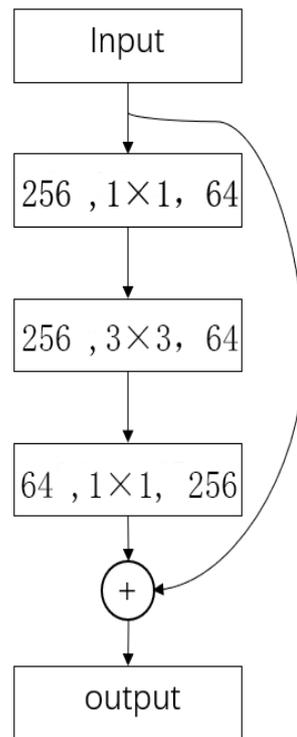


Figure 1. Block of ResNet.

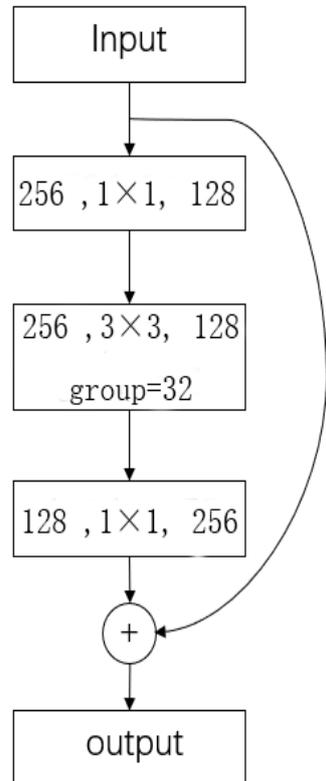


Figure 2. Block of ResNeXt.

3. RFSE-ResNeXt Model Structure

3.1. Squeeze Excitation Structure

HU et al. proposed the squeeze–excitation (SE) structure from the channel dimension in 2017 [16]. The core idea of the SE module is to obtain the weights of each feature channel while the network is being trained and to make the weights of the feature maps that are favorable to the loss function decrease more, while the weights of the feature maps that do not contribute or contribute little to the loss function decrease less. Thus, the effect of boosting useful features and suppressing useless features is achieved. The specific structure is shown in Figure 3.

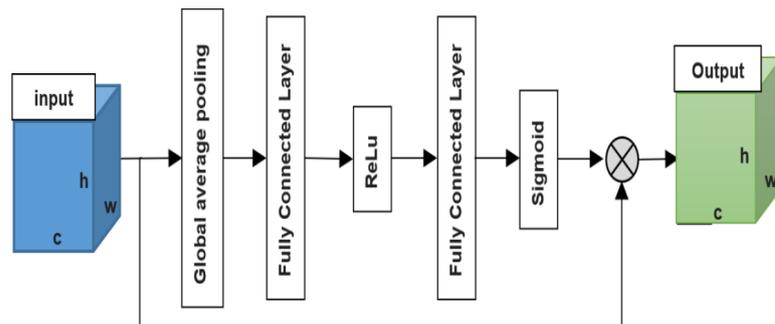


Figure 3. Squeeze and excitation structure.

The SE structure is divided into a squeeze step and an excitation step.

In the squeeze step, the original feature map $U \in R^{H \times W \times C}$ (H, W, C represent the height, width, and number of channels, respectively) is convolved and pooled to obtain a global average pooling (GAP) with the output dimension of $1 \times 1 \times C$.

In the excitation step, the globally averaged features obtained in the squeeze step are fed into a two-level full connection (FC), a rectified linear unit (ReLU) [17], and a Sigmoid function [18] to output a channel weight vector of dimension $1 \times 1 \times C'$ ($C' < C$) of the channel weight vector. where the output layer changes from $1 \times 1 \times C'$ to $1 \times 1 \times C$, i.e., the weights assigned to the features represented by each channel.

Finally, the output is multiplied by the original feature map U to obtain a feature map incorporating the channel attention information $Z \in R^{H \times W \times C}$.

3.2. Residual Fusion Structure

The convolutional network used in this paper is ResNeXt101 [19], and the main body consists of four residual modules. The structure of the residual modules is shown in Figure 4. In the input–output convolution layer, a small convolutional kernel of $1 \times 1 \times 1$ is used, and the middle group convolutional structure uses the usual base 32 with a size of $3 \times 3 \times 3$. The advantage of ResNeXt101 is that the group convolution is performed by stacking blocks of the same topology in one parallel, which improves the accuracy without increasing the parameter complexity.

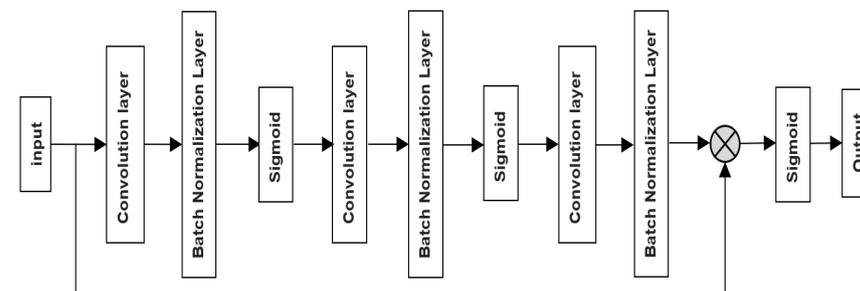


Figure 4. Residual module.

The improved residual structure is shown in Figure 5. Where layer1, layer2, layer3, and layer4 denote the four residual modules in the network, respectively. In the original network, the deepening of layers will make some detailed features filtered out, resulting in insufficient utilization of the extracted features. In this paper, the improved residual module structure of the ResNeXt network makes full use of the feature information extracted from each layer by fusing the detailed features extracted from the shallow network with the features extracted from the deep network.

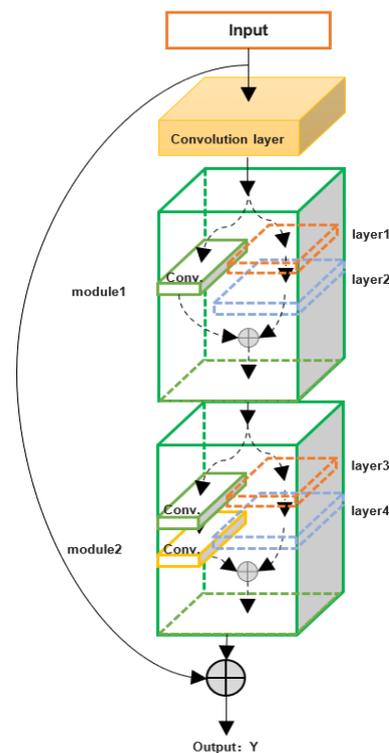


Figure 5. Improved ResNeXt residual structure.

First, the original data input is convolved with a convolution kernel size of 1×1 , and the features obtained after convolution are fused with the output of the features from layer2 as the output of module1. Then, the features output from module1 are convolved twice with a kernel size of 1×1 and then fused with the features output from layer4. The effect of mutual fusion of the signal features extracted by the residual module is achieved.

The convolution operation is performed first to reduce the feature dimensionality and enable the features to be fused. In this paper, layer-hopping fusion is used instead of the layer-by-layer feature fusion. Firstly, because ResNeXt101 itself has a considerable number of parameters, if layer-by-layer fusion is used, the overall computation will be too large, and the training time will be too long. Secondly, if the layer-by-layer fusion is used, some redundant information will be fused and overfitting will be unavoidable, which will instead reduce the recognition accuracy of the network. The two features are fused using element-wise [20].

3.3. Establishing the RFSE-ResNext Network

In this paper, the residual module in ResNeXt101 [19] is first modified as described above, and then the squeeze–excitation structure is introduced to obtain the RFSE-ResNeXt network. Figure 6 shows the core junction RFSE-ResNeXt block of the network. The original input of the signal is first passed through the modified residual fusion structure to fully extract the shallow and deep features. The output is then multiplied with the feature

information processed by the SE structure to enhance the useful features and suppress the interference features. Each of these convolutional operations consists of convolution, batch normalization (BN) [21], and ReLU.

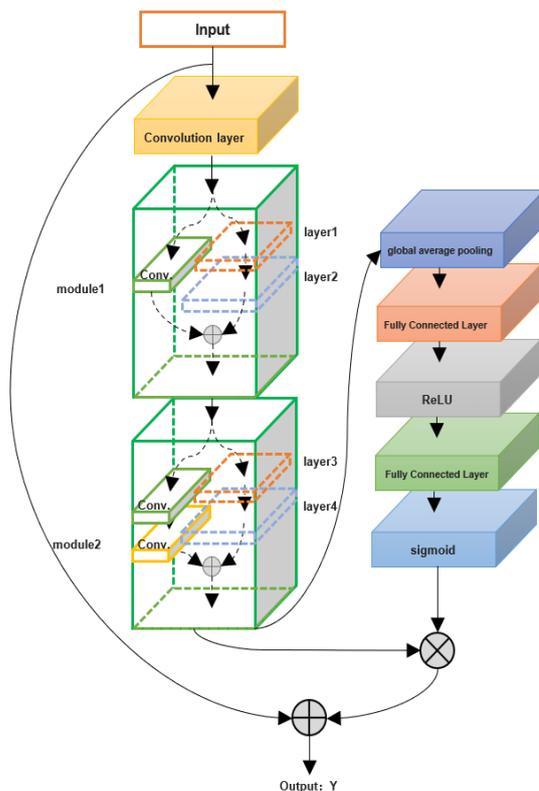


Figure 6. Block structure of RFSE-ResNeXt.

The final RFSE-ResNeXt network consists of one convolutional layer, four RFSE-ResNeXt block structures, two pooling layers, one fully-connected layer, and one Softmax classifier, and the network structure is shown in Figure 7. The network input layer determines the input data as IQ bidirectional time-domain signals with each sample data of size 2×128 . The raw data is first input to the first convolutional layer (consisting of 50 convolutional kernels of size 2×8), and after the four RFSE-ResNeXt block structures and the global average pooling, the final classification is performed using the Softmax classifier.

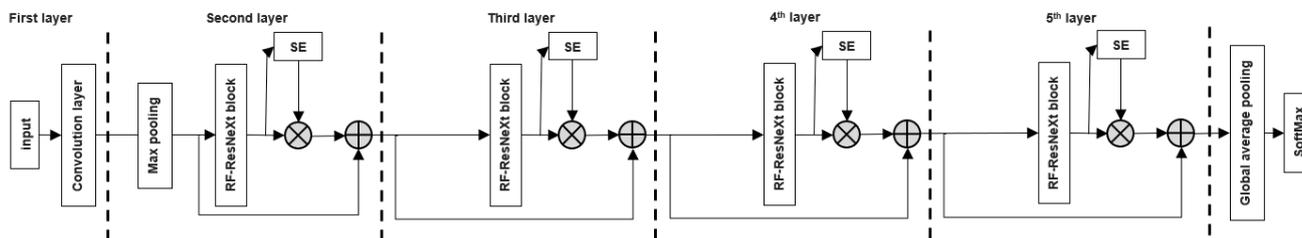


Figure 7. RFSE-ResNeXt network structure.

4. Analysis of Model Results

4.1. Datasets

In this paper, we use the international common dataset RML2016.10a [22] for the study. This dataset contains 8 classes of digital modulated signals (BPSK, 8PSK, CPFSK, GFSK, PAM4, QAM16, QAM64, QPSK) and 3 classes of analog modulated signals (AM-DSB, AM-SSB, WBFM), a total of 11 different modulated signals.

The signal-to-noise ratios of the 11 signals are distributed in the range of -20 to 18 dB with an interval of 2 dB. Each signal consists of two IQ channels with 128 sampling points at different signal-to-noise ratios. In addition, this dataset adds the influence factors such as central frequency shift, fading out, and additive Gaussian white noise to various signals in order to closely match the real environment.

For the comparative analysis of network performance, a total of $221,000$ samples from the RML2016.10a dataset were used, of which 70% were used as the training set and 30% as the test set. All network models in this paper use theano back end and Keras framework. The hardware configuration is CPU Intel i7-8750CPU, GPU Nvidia GeForce RTX2080.

4.2. Impact of Residual Fusion Structure on Network Performance

As shown in Figure 8, the improved ResNeXt network structure has a fairly good recognition accuracy overall. Although there is a slight decrease in accuracy at low signal-to-noise ratios, there is an improvement in both the middle and high segments. The reason is that the group convolution strategy of the ResNeXt network itself improves the recognition accuracy but reduces the generalization performance (i.e., the effect is reduced when facing complex signals with low signal-to-noise ratio). In contrast, in this paper, the residual structure of ResNeXt is improved so that the extracted deep and shallow features are fused with each other. Although this improvement suggests the overall performance of the network, it has little effect in compensating for the drawback of generalization performance. Therefore, we introduce a compressed excitation structure on top of the residual fusion structure to improve the generalization ability of the network.

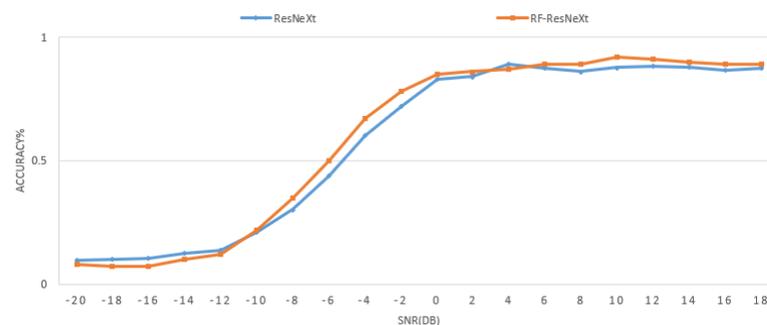


Figure 8. Performance comparison before and after improvements.

4.3. Impact of Squeeze–Excitation Structure on Network Performance

In order to investigate the effect of the incorporated squeeze–excitation structure on the classification ability of the network, we will compare the models with and without the incorporated squeeze–excitation structure by using the parameters of global accuracy and -20 dB to 0 dB average accuracy of the network model on the test set as the evaluation criteria. The specific data are shown in Table 1.

Table 1. Effect of adding a squeeze–excitation structure on the model.

Models	Global Accuracy (%)	-20 dB to 0 dB Average Accuracy (%)
RF-ResNeXt	59.1	33.2
RFSE-ResNeXt	59.8	34.6

As can be seen from the table, the global accuracy of the model is only slightly improved after adding the squeeze–excitation structure, but the average accuracy at low signal-to-noise ratios is significantly improved. The reason is that the squeeze–excitation module provides an effective solution to the information overload problem that exists in the ResNeXt model. The SE module makes the weights of the feature maps that contribute

to the loss function decline larger by obtaining the weights of each feature channel, while the weights of the feature maps that do not contribute or contribute less to the loss function decline become smaller. Thus, the accuracy and efficiency of the network are improved by boosting useful features and suppressing useless features.

4.4. Performance Analysis of Different Networks

The experimental analysis compares the RFSE-ResNeXt network, an optimized CLDNN network in the literature [7], an improved Resnet network in the literature [23], a CLDNN network, a vgg-cnn network, and an inception network in this paper. The effectiveness of various network models for the classification task is shown in Figure 9.

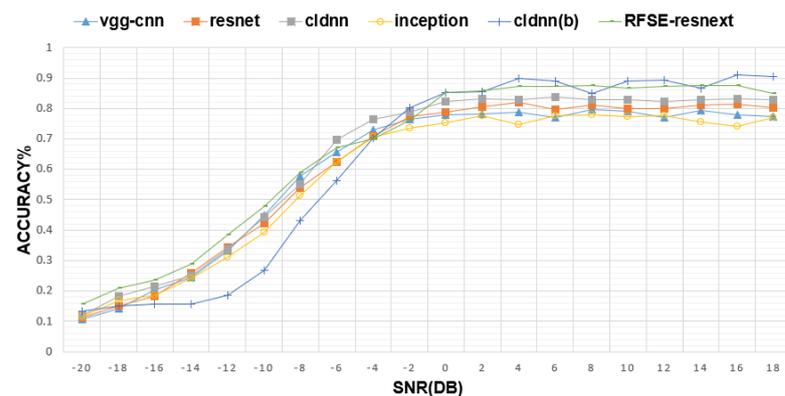


Figure 9. Comparison of the recognition performance of six models (RML2016.10A).

As shown in Figure 9, the model proposed in this paper shows a considerable improvement in recognition rate in the -20 dB to -8 dB interval even when compared with the current newer models and also has a fairly good recognition accuracy at high signal-to-noise ratios. The rFSE-ResNeXt can achieve this result, and it is important because rFSE-ResNeXt is the only model among these that has the SE module and advanced grouped convolutional structure. The former allows the network to use more energy to focus on the typical features of the signal and achieve better results with the same number of parameters; the latter allows the network to extract sample features significantly, further improving the accuracy of the network. Meanwhile, RFSE-ResNeXt further improves the ability of extracting signal features due to the improved residual network structure, which makes the deep and shallow features fuse. Because of this, it has quite good performance in the face of the close-to-reality RML2016.10A dataset used in this paper.

Figures 10 and 11 show the confusion matrix plots of ResNeXt and the network model in this paper at the signal-to-noise ratio of 0 dB, respectively, which can visually reflect the recognition effect of the model for the 11 signal modulation methods. Since ResNeXt adopts the strategy of group convolution, although it greatly improves the ability of extracting feature information, it does not perform well in the face of signals with low signal-to-noise. It can be seen that there is a problem of misidentifying some WBFM as AM-DSB at 0 dB, and confusing some 8PSK and QPSK, which leads to a low recognition rate of 8PSK, WBFM, and QPSK signals. The overall recognition accuracy of the RFSE-ResNeXt network model proposed in this paper is improved, especially since the phenomenon of confusing AM-DSB into WBFM is attenuated.

In the face of a very low signal-to-noise ratio of -10 dB, ResNeX can only identify almost one modulated signal of QAM64, and all kinds of signals are highly confused into CPFSK. while the proposed model in this paper has considerable attenuation of the confusion of AM-DSB, BPSK, BPSK, etc., which improves the overall recognition accuracy. The reason for this is that the residual fusion structure fully extracts the deep features of the signals at low signal-to-noise ratios, while the squeeze–excitation structure enables the network to learn more about the key information for correct classification. Table 2 shows

the confusion between the network in this paper and the ResNeXt network for CPFSK signals at a signal-to-noise ratio of -10 dB.

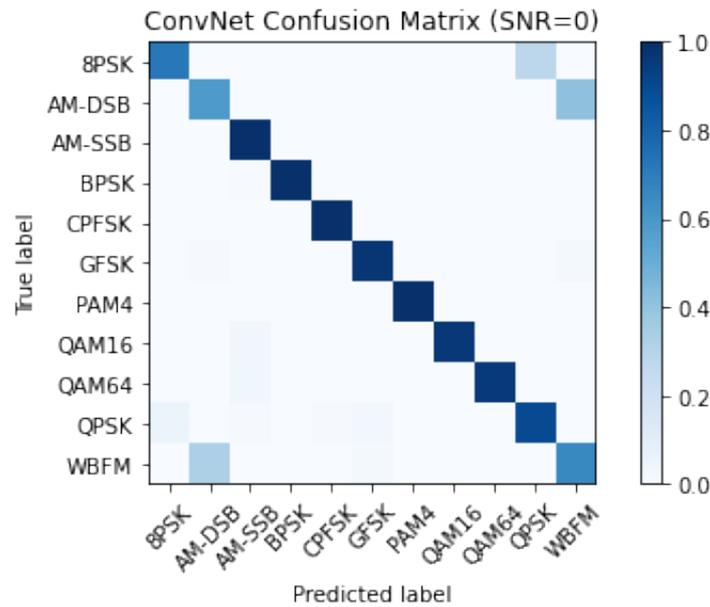


Figure 10. Confusion matrix on SNR = 0 dB (ResNeXt).

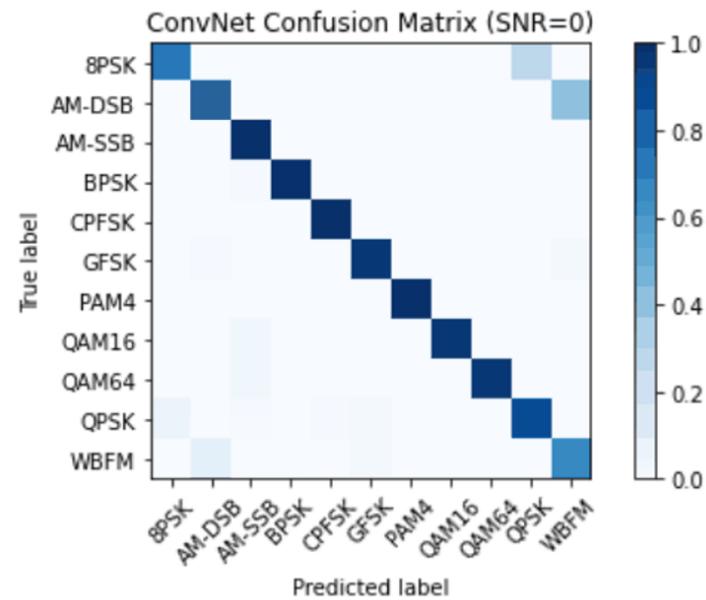


Figure 11. Confusion matrix on SNR = 0 dB (RFSE-ResNeXt).

Table 2. CPFSK recognition at SNB = -10 dB.

ResNeXt	RFSE-ResNeXt
About 28% AM-DSB misclassification into CPFSK	About 12% AM-DSB misclassification into CPFSK
About 56% BPSK misclassification into CPFSK	About 27% BPSK misclassification into CPFSK
About 45% of GFSK misclassified into CPFSK	About 31% of GFSK misclassified into CPFSK

5. Conclusions

In this paper, the RFSE-ResNeXt network is established based on the ResNeXt network. By improving the residual structure of the ResNeXt network, the deep and shallow features are fused, and the feature information extracted from each level is fully utilized. Then the squeeze–excitation structure is introduced to improve the generalization ability of the network so that the recognition accuracy of the network can be improved when the complex signal-to-noise ratio is low. The experimental results show that when the signal-to-noise ratio is as low as -8 dB or less, the network has a better adaptation ability to low signal-to-noise ratios and stronger anti-confusion ability compared with existing methods.

Author Contributions: Conceptualization, X.Z. and B.Z.; methodology, B.Z. and H.Y.; software, F.W.; formal analysis, F.W.; validation, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Available online: <https://www.Deepsig.io/datasets>, accessed on 27 July 2022.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ResNeXt	Aggregated Residual for Networks
SE	Squeeze-Excitation
RF	Residual-Fusion
RF-ResNeXt	Residual-fusion aggregated residual for networks
RFSE-ResNeXt	Residual-fusion squeeze–excitation aggregated residual for networks

References

1. Wang, S.Q.; Bai, J.; Huang, X.Y.; Gao, C.Y.; Wan, P.F. Analysis of radar emitter signal sorting and recognition Model structure. *Pro-Cedia Comput. Sci.* **2019**, *154*, 500–503. [CrossRef]
2. Wang, C.; Wang, J.; Zhang, X.D. Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 5–9 March 2017; pp. 2437–2441.
3. Wang, Y.; Zhang, S.; Zhang, Y.; Wan, P.; Li, J.; Li, N. Acooperative spectrum sensing method based on Empirical mode composition and information geometry in complex electromagnetic environment. *Complexity* **2019**, *2019*, 5470974.
4. Zhang, M.; Diao, M.; Gao, L.; Liu, L. Neural networks for radar waveform recognition. *Symmetry* **2017**, *9*, 75. [CrossRef]
5. Chen, T.; Liu, L.Z.; Huang, X. LPI radar waveform recognition based on multi-branch MWC compressed sampling receiver. *IEEE Access* **2018**, *6*, 30342–30354. [CrossRef]
6. Zhou, Z.W.; Huang, G.M.; Chen, H.Y.; Gao, J. Automatic radar waveform recognition based on deep convolutional denoising autoencoders. *Circuits Syst. Signal Process.* **2018**, *37*, 4034–4048. [CrossRef]
7. Zou, B.; Zeng, X.; Wang, F. Research on Modulation Signal Recognition Based on CLDNN Network. *Electronics* **2022**, *11*, 1379. [CrossRef]
8. O’Shea, T.J.; Corgan, J.; Clancy, T.C. Convolutional radio modulation recognition networks. In Proceedings of the International Conference on Engineering Applications of Neural Networks, Aberdeen, UK, 2–5 September 2016; Springer: Berlin/Heidelberg, Germany, 2016.
9. West, N.E.; O’Shea, T.J. Deep architectures for modulation recognition. *arXiv* **2017**, arXiv:1703.09197.
10. O’Shea, T.; West, N. Radio machine learning dataset generation with gnu radio. In Proceedings of the GNU Radio Conference, Boulder, CO, USA, 12–16 September 2016.
11. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. *AAAI* **2017**, *4*, 12. [CrossRef]
12. O’Shea, T.J.; Roy, T.; Clancy, T.C. Over the Air Deep Learning Based Radio Signal Classification. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 168–179. [CrossRef]
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

14. Xie, S.; Girshick, R. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
15. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Wey, T. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
16. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
17. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. In *International Conference on International Conference on Machine Learning*; Omnipress: Madison, WI, USA, 2010.
18. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; JMLR Workshop and Conference Proceedings; pp. 315–323.
19. Mahajan, D.; Girshick, R.; Ramanathan, V.; He, K.; Paluri, M.; Li, Y.; Van Der Maaten, L. Exploring the Limits of Weakly Supervised Pretraining. In *Computer Vision-ECCV*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11206.
20. Lee, J.; Kim, D.; Ham, B. Network Quantization with Element-wise Gradient Scaling. *arXiv* **2021**, arXiv:2104.00903.
21. Lofe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
22. O’Shea, T.J. Available online: <https://www.Deepsig.io/datasets> (accessed on 8 January 2022).
23. Qin, X.; Huang, J.; Zha, X.; Luo, L.P.; Hu, D.X. Radar emitter signal recognition based on dilated residual network. *Acta Electronica Sin.* **2020**, *48*, 456–462.