GCPS: A CNN Performance Evaluation Criterion for Radar Signal Intrapulse Modulation Recognition

Zhengyang Yu [ORCID], Jianlong Tang, and Zhao Wang

Abstract—In radar intrapulse modulation recognition through convolutional neural networks (CNN), the recognition rate vis-a-vis varying signal-to-noise ratios (SNRs) is utilized to analyze the performance of networks. However, three shortcomings exist in this realm, (1) Proportional increase in recognition rate up to 100%, due to SNR increase, leading to saturation. (2) The dependency of recognition rate on the test set. (3) Recognition rate failing to substantiate the increase in SNR and its manifestation in CNN performance enhancement. Foregone in view, this letter proposes an evaluation criterion to evaluate CNN performance to address the shortcomings above. We propose Grad-CAM position score-Internal Benchmark (GCPS-IB) to address (1), (2), and Grad-CAM position score-External Benchmark (GCPS-EB) to address (3). Finally, we utilize GCPS to evaluate and compare the performance of existing GoogLeNet and ResNet-18 to verify the efficacy. The proposed criterion establishes its efficacy by addressing the shortcomings mentioned above, hence augmenting the recognition rate.

Index Terms—Convolutional neural network (CNN), gradient-weighted class activation mapping (Grad-CAM), radar signal processing, time-frequency image.

I. INTRODUCTION

RADAR signal intrapulse modulation recognition is one of the critical technologies of electronic intelligence (ELINT) in electronic warfare (EW) [1]. This technology can help us identify unknown radar signals and respond quickly. However, with the development of a low probability of intercept (LPI) radar, the reliability of traditional recognition methods that rely on radar signal parameters has substantially decreased [2]. In recent years, due to the rapid development of deep learning, converting radar signals into images and using convolutional neural networks (CNNs) for recognition has become the mainstream method [3].

With the continuous evolution of the CNN structure, the recognition rate of radar signal intrapulse modulation has also increased. Although increasing the number of layers can improve the recognition rate of CNNs, it is not easy to interpret and evaluate such a large CNN structure [4]. At present, researchers mainly study the explanation of CNNs from the aspect of feature visualization. Quanshi Zhang expressed the CNN internal logic by splitting convolutional layer features, and proposed a knowledge graph to find the feature extraction area of CNN. Through the class activation mapping (CAM)

and the gradient-weighted CAM (Grad-CAM) [5], users can more clearly understand how CNN makes recognition decisions. However, CNN performance evaluation is still qualitative and intuitive, and precise quantitative mathematical tools have not yet been developed.

Therefore, a quantitative performance evaluation method for CNN performance in conventional optical image recognition (i.e., images containing items, people, handwriting, etc.) is still challenging [4]. The current strategy in radar signal intrapulse modulation recognition is to utilize only the existing CNN structure to recognize a radar signal time-frequency image (TFI) [3] without analyzing the reason for the change in recognition rate. Consequently, evaluating the CNN performance for radar signal intrapulse modulation recognition is needed.

Although CNNs are challenging to describe directly through mathematical formulas [6], if a measurement of the CNN performance difference for different samples can be developed, an indirect quantitative evaluation of CNNs can be achieved. Compared with conventional images, radar signal TFIs have superficial mode characteristics; complete, smooth boundaries; and distinct signal and no-signal regions. Considering these specific attributes, we can quickly obtain the radar signal's position, and it is possible to evaluate the performance of the CNN and analyze the influence of noise quantitatively.

To achieve these goals, we propose the Grad-CAM position score (GCPS), which consists of the GCPS-Internal Benchmark (GCPS-IB) and GCPS-External Benchmark (GCPS-EB). The benchmark is set up based on a Grad-CAM radar signal TFI without noise. GCPS-IB and GCPS-EB quantitatively evaluate the CNN performance and the influence of noise on feature extraction by the CNN, respectively, in radar signal intrapulse modulation recognition. Additionally, we use these indicators to evaluate existing GoogLeNet and ResNet-18 models, and the experimental results verify that the criterion we propose is valid.

II. THE PROPOSED METHOD

A. Architecture Overview

Fig. 1 shows the method architecture for obtaining the GCPS. The white part is the radar signal construction step.

The received radar signal $s(t)$ is

$$s(t) = x(t) + n(t) \tag{1}$$

where $x(t)$ denotes a radar signal, and $n(t)$ is Gaussian noise. The radar signal can be expressed as an analytical signal, as shown in the following formula:

$$s(t) = A(t) \exp\left[j\left(2\pi f_0 t + c(t) + \varphi_0\right)\right] \tag{2}$$

where $A(t)$ and f_0 are the signal amplitude and carrier frequency, respectively. $c(t)$ is the phase function, which determines the modulation type of the radar signal. φ_0 is the initial phase.

The two indicators of GCPS generation can be divided into three parts as follows.

1) Signal processing and identification (red parts).

2) The benchmark, objective, and interference definitions (coral and blue parts).

3) GCPS-IB and GCPS-EB calculation (yellow parts).

In this letter, we constructed radar signals of five modulation types as training samples for CNNs: single-carrier (SC) modulation, frequency-shift keying (FSK), linear frequency modulation (LFM), parabolic frequency modulation (PFM), and sine frequency modulation (SFM). Gaussian white noise is introduced and controls the noise intensity to change the signal-to-noise ratio (SNR).
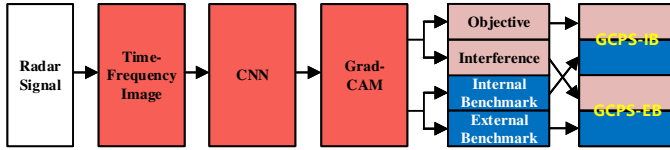

Fig. 1. Architecture of our method, which consists of radar signal construction, signal processing and identification, GCPS parameter generation, and GCPS calculation.

B. Signal Processing and Identification

This part consists of three steps: TFI generation, identification via a CNN, and Grad-CAM processing.

1) TFI Generation

Using a CNN for identification, we need to convert the radar signals to an image. The time-frequency distribution is the conventional algorithm to obtain the TFI of the radar signal. We selected the short-time Fourier transform (STFT) as the signal TFI extraction algorithm [7]. The formula of STFT is as follows:

$$S(t,f) = \int_{-\infty}^{+\infty} s(\tau)h(\tau-t)e^{-j2\pi f\tau}d\tau \qquad (3)$$

where $h(t)$ is the window function and $S(t, f)$ is the signal's frequency spectrum. By sliding $h(t)$, the Fourier transform is reliably applied to the signal, segment by segment and obtains the signal's time-frequency matrix. Therefore, we can transform the time-frequency matrix into a TFI. Fig. 2 shows the STFT results of the five modulation types of radar signals.

The signal boundary of each TFI is complete and smooth since there is no noise. Under the influence of noise, the signal boundary becomes susceptible to irregularity and flutter, which may affect CNN feature extraction performance.

2) Identification via a CNN

GoogLeNet won 1st place in the ImageNet localization task in the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [8] and reached 6.67% in the top-5 error rate. The number of parameters were also less than half of AlexNet. Therefore, GoogLeNet is an excellent and practical CNN structure. K. He proposed ResNet and won 1st place in

the same competition as GoogLeNet in ILSVRC 2015 [9]. The top-5 error rates were as low as 3.57% compared to GoogLeNet, considering the experimental conditions. Because of its good performance and wide use, we chose GoogLeNet and ResNet-18 as the standard CNNs. Moreover, experiments also verify the ability of the GCPS to evaluate different CNNs.

(a) SC (b) FSK (c) LFM (d) PFM (e) SFM

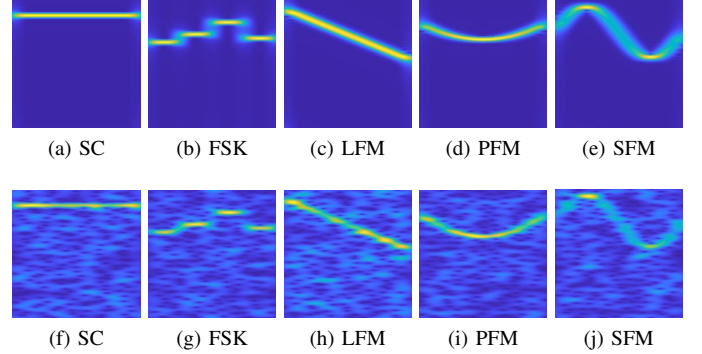(f) SC (g) FSK (h) LFM (i) PFM (j) SFM

Fig. 2. TFIs of radar signals with five modulation types. (a, b, c, d, and e) are SC, FSK, LFM, PFM, and SFM TFIs without noise. (f, g, h, i, and j) are the same as (a, b, c, d, and e), respectively, when the SNR is equal to 0dB. The sampling point of each signal is 224.

3) Grad-CAM Processing

Grad-CAM uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localized map highlighting the crucial regions in the image for predicting the concept [5]. Based on this algorithm, we propose a criterion to evaluate CNN performance and noise influence quantitatively. Taking ResNet-18 as an example, Fig. 3 shows the Grad-CAM radar TFIs.
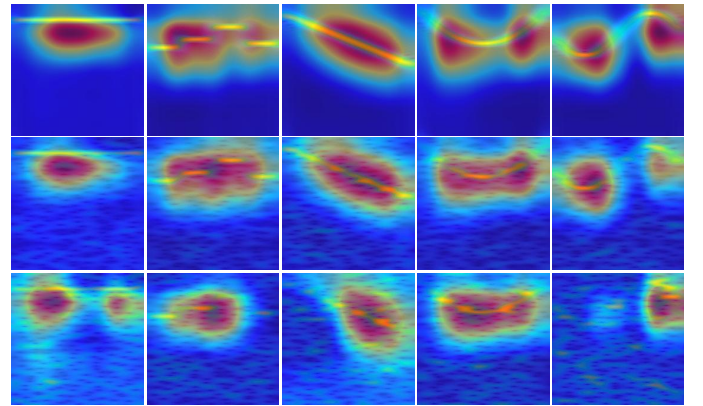

Fig. 3. ResNet-18 Grad-CAM radar TFIs for different SNRs. The Grad-CAM radar TFIs of the five modulation types without noise are in the first row. The SNR in the second row is equal to 0dB. The SNR in the last row is equal to -6dB.

Without noise, the red region (high score for the class) and the radar signal region have a high degree of coincidence. As the SNR decreases, the red region gradually deforms and breaks away from the signal region, which means that the degree of noise influence increases when the CNN extracts features.

C. Benchmark, Objective, and Interference Definition

To obtain the GCPS, we first need to calculate the benchmark, objective, and interference. In the following, the TFI of

LFM is used to introduce each parameter.

1) Benchmark

The Grad-CAM results reveal the CNN sensitive region of the TFI, which is the position that the CNN used for classification. Therefore, we treat the red areas in the Grad-CAM TFI without noise as the internal benchmark (IB) and the other areas as the external benchmark (EB).

The IB represents the position where the CNN should extract features. The EB represents the incorrect area, which may interfere with classification if the CNN extracts feature in this region. The IB and EB are defined as follows:

$$\text{IB} = \{(i,j) \mid \text{GT}(i,j) \geq Rth, i \leq \text{W}, j \leq \text{H}, i,j \in N^*\} \quad (4)$$

$$\text{EB} = \complement_U \text{IB} \quad (5)$$

where GT is the Grad-CAM TFI without noise, the TFI is in the RGB color space, and Rth is the red channel threshold of the TFI and is set to 250. W and H are the width and height respectively, of the TFI. N^* is the natural number set. The EB can be obtained simply by calculating the complement of the IB.

2) Objective and Interference

As shown in Fig. 4(b), the number of red points within the IB begins to decrease as the SNR decreases. The red points in the Grad-CAM TFI with noise and within the IB are called the objective. The objective is defined as follows:

$$\text{Objective} = \{(i,j) \mid \text{GTN}(i,j) \geq Rth, (i,j) \in \text{IB}\} \quad (6)$$

where GTN is the Grad-CAM TFI with noise.

From Fig. 4(c), due to the influence of noise, CNN feature extraction areas gradually shift to the noise region. The red points in the IB almost disappear and appear in the EB. In this situation, we name the red points as interference. The interference is given by

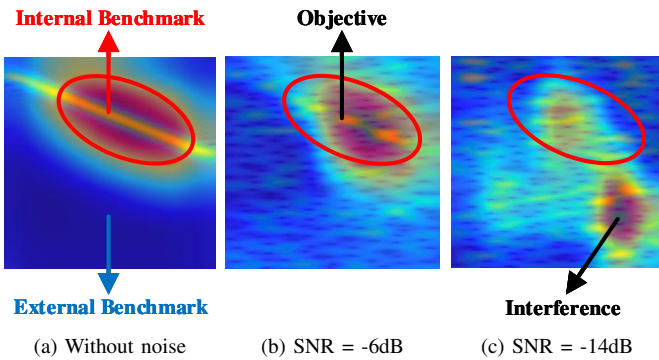$$\text{Interference} = \{(i,j) \mid \text{GTN}(i,j) Rth, (i,j) \in \text{EB}\} \quad (7)$$

Fig. 4. Schematic diagram of the benchmark, objective, and interference. (a) is radar signal TFI without noise. (b) and (c) are TFIs with noise.

D. GCPS-IB and GCPS-EB Calculation

When obtaining the four sets (i.e., the IB, EB, objective, and interference), the GCPS-IB and GCPS-EB are defined as follows:

$$\text{GCPS-IB} = \frac{\mid \text{Objective} \mid}{\mid \text{IB} \mid} \quad \text{GCPS-IB} \in [0,1] \quad (8)$$

$$\text{GCPS-EB} = \frac{\mid \text{Interference} \mid}{\mid \text{EB} \mid} \quad \text{GCPS-EB} \in [0,1] \quad (9)$$

where $\mid$ IB $\mid$, $\mid$ EB $\mid$, $\mid$ Objective $\mid$, and $\mid$ Interference $\mid$ are the total number of elements in the respective sets. According to formulas (6) and (7), the ranges of GCPS-IB and GCPS-EB are both [0, 1].

The GCPS-IB reveals the CNN's ability to extract correct features, and the GCPS-EB reveals the extent to which the CNN extracts incorrect features because of noise interference. GCPS-IB and GCPS-EB constitute GCPS, which supplements the recognition rate to evaluate CNN's performance quantitatively.

To facilitate analysis of noise influence on the criterion and recognition rate, other factors were not taken into consideration, which may affect the recognition rate. For example, LPI radar signals are not easy to detect [3].

III. EXPERIMENT

A. Dataset Introduction

There were five modulation types of radar signals in the experiments: SC, FSK, LFM, PFM, and SFM. We adjust the SNR by adding Gaussian noise. Additionally, to verify the generalizability of the GCPS, the parameters of each radar signal changed randomly within the specific ranges. Besides, the sampling point for each signal is 224. The table below indicates the radar signal parameter setting, and the frequencies in the table are normalized, which means that if the sampling frequency (f_s) is 1 GHz, the range of the SC modulation is $[0.1 \times f_s, 0.4 \times f_s] = [100\text{MHz}, 400\text{MHz}]$.

TABLE I
PARAMETER SETTING OF RADAR SIGNAL

Type	Parameter	Ranges
SC	Carrier frequency f_0	$0.1 \sim 0.4$
FSK	$f_1 \sim f_4$	$0.1 \sim 0.4$
LFM	f_{min}	$0.01 \sim 0.35$
	Bandwidth B	$0.15 \sim 0.4$
PFM	f_{min}	$0.01 \sim 0.4$
	B	$0.1 \sim 0.4$
SFM	f_{min}	$0.01 \sim 0.3$
	B	$0.2 \sim 0.4$

12,500 radar signals constitute the data set and are evenly distributed among five modulation types and ten signal-to-noise ratios. ([-4dB, 14dB], in 2dB steps). We used 80% percent of the dataset for training and 20% for validation. The SNR of the test set was [-14dB, 14dB], in 2dB steps, and the total number of samples in the test set was $15 \times 5 \times 100 = 7500$. After completing the sample construction, we selected GoogLeNet and ResNet-18 for experiments.

B. Comparison Based on Different Radar Signals

As shown in Fig. 5(a) and Fig. 5(b), the recognition rate of GoogLeNet and ResNet-18 increased rapidly when 14dB

$\leq$ SNR $<$ -5dB and gradually reached 100% when -5dB $\leq$ SNR $\leq$ 0dB. When SNR $>$ 0dB, both CNN recognition rates on the five modulation types of radar signals reached 100%.

Fig. 5(c) and Fig. 5(d) show that the GCPS-IB had a similar trend as the recognition rate. More importantly, in contrast to the scenario where the recognition rate always equals 100% when SNR $>$ 0dB, a GCPS-IB distinction was made over the entire range of SNR. The recognition rate is the test stage's statistical value, which can only be used as a reference when identifying new samples. However, the generation of GCPS-IB does not have this restriction. Therefore, GCPS-IB is an excellent indicator to evaluate CNN performance in radar signal recognition quantitatively.

Fig. 5(e) and Fig. 5(f) show the GCPS-EB trend for different SNRs. Because the signal processing part may introduce unknown interference, we found that the GCPS-EB of GoogLeNet suddenly dropped when SNR $<$ -2dB. Except for this condition, the experimental results show that the GCPS-EB is inversely proportional to the SNR, which means that it can quantitatively evaluate the effect of noise on the CNN feature extraction. Comparing Fig. 5(a, c, e) and Fig. 5(b, d, f), ResNet-18 is more robust than GoogLeNet since the recognition rate trend of the five types of radar signals were closer to each other.
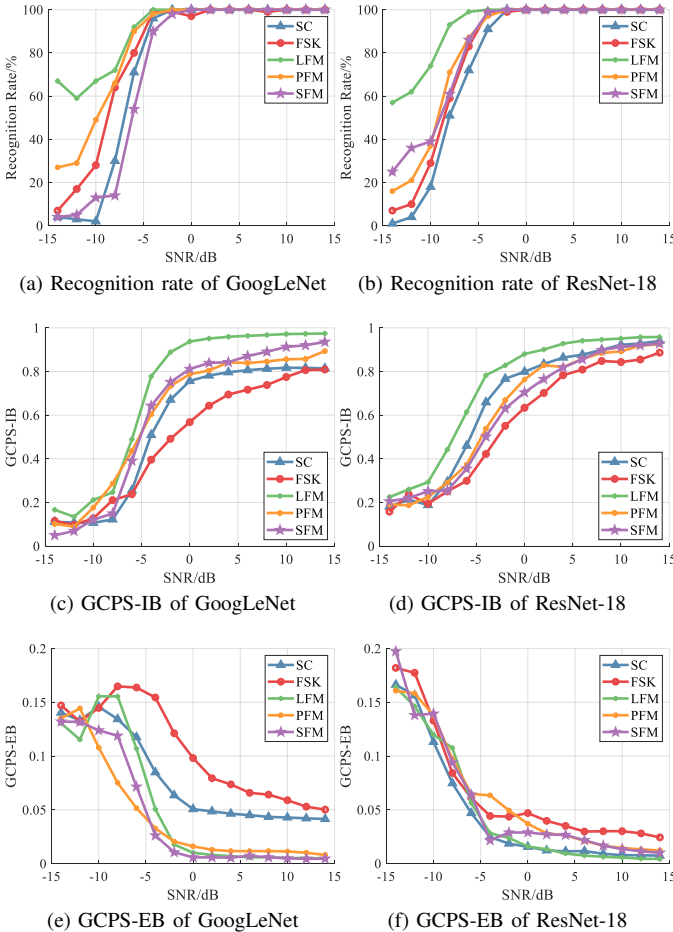
C. Comparison Based on Different CNN

The above experiments confirmed that GCPS-IB and GCPS-EB are reliable and effective. To further explore how the GCPS evaluates the performance of different CNNs and verify the relationship between the GCPS and recognition rate, we performed an overall analysis of different radar signals' experimental results. We implemented them in the same coordinate system. Fig. 6 shows the comparable results of the two CNNs.

(a) GCPS-IB comparison (b) GCPS-EB comparison

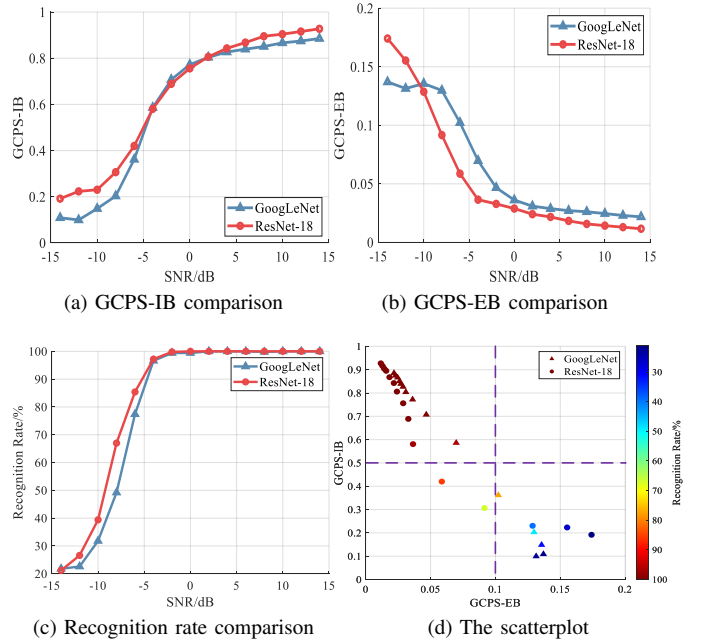(c) Recognition rate comparison (d) The scatterplot

Fig. 6. GoogLeNet and ResNet-18 performance evaluation and the relationship between the GCPS and recognition rate. Additionally, (d) describes the relationship between the GCPS and the recognition rate.

The GCPS was used to evaluate the performance of GoogLeNet and ResNet-18. Higher GCPS-IB and GCPS-EB scores, lead to extraction of more objective and interference features, respectively. As shown in Fig. 6(a) and Fig. 6(b), the GCPS-IB of ResNet-18 is higher than that of GoogLeNet at different SNRs. GCPS-EB of ResNet-18 is lower than that of GoogLeNet when SNR $\geq$ -10dB. This finding means that ResNet-18 can rely on more objective features for recognition. Therefore, we consider that the performance of ResNet-18 is better than that of GoogLeNet without using the test set to obtain the CNN recognition rate for judgment.

From Fig. 6(c), the recognition rate of ResNet-18 is higher than that of GoogLeNet when SNR $\leq$ 0dB. When SNR $>$ 0dB, the recognition rates of ResNet-18 and GoogLeNet both reached 100%. We can conclude that the performance of ResNet-18 is better than that of GoogLeNet. However, the current signal environment is complex and variable [3]. It is not easy to control types, parameters, and quantity of the acquired radar signals, as in experiments. Suppose these insufficient radar signals are used as a test set to determine CNN's recognition rate, this recognition rate's reliability will become very low. It will be unreliable to compare the performance of the network by only the recognition rate.

(a) Recognition rate of GoogLeNet (b) Recognition rate of ResNet-18

(c) GCPS-IB of GoogLeNet (d) GCPS-IB of ResNet-18

(e) GCPS-EB of GoogLeNet (f) GCPS-EB of ResNet-18

Fig. 5. Experimental results of the radar signals of the five modulation types for different SNRs.

We also analyzed the relationship between the GCPS and recognition rate based on GoogLeNet and ResNet-18.

Fig. 6(d) is the scatterplot in which the horizontal and vertical coordinates represent the GCPS-EB and GCPS-IB, respectively. The colors of the scattered points represent the recognition rates. Moreover, we use the median value of the GCPS-IB and GCPS-EB to divide the image into four quadrants for facilitating analysis.

From Fig. 6(d), very few scattered points are located in the first or third quadrant because the GCPS-IB is inversely proportional to the GCPS-EB. Most of the scattered points were in the second quadrant, when the recognition rate of the two networks was close to 100%. We can conclude that the recognition rate is proportional to the GCPS-IB and inversely proportional to the GCPS-EB.

The GCPS-IB of ResNet-18 was higher than that of GoogLeNet and the GCPS-EB of ResNet-18 was lower than GoogLeNet under the same recognition rate. It means that ResNet-18 has a higher probability of extracting features from the objective region instead of the interference region and ResNet-18 has a strong anti-noise ability. Therefore, ResNet-18 outperforms GoogLeNet.

D. Why the GCPS Works

We use six confusion matrices to show how the GCPS can represent the recognition reliability. The reliability here is not a degree of confidence but refers to the possibility of an indicator error [10]. It means that the higher the reliability, the more radar signals CNN can recognize correctly.
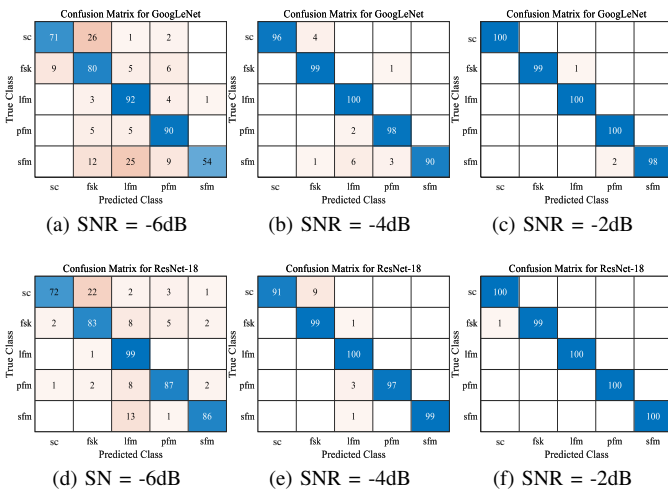

Fig. 7. Recognition results of GoogLeNet and ResNet-18 under different SNRs.

The recognition rate is a statistic and will be affected by the selection of the test set. It is challenging to evaluate CNN's overall performance only through Fig. 5(a) and Fig. 5(b). Simultaneously, CNN has different recognition effects for various types, but counting the recognition rates as Fig. 6(c) will lose this information. Losing information will affect the reliability of the recognition rate. The key to the usefulness of GCPS is that it can combine the CNN recognition results for different types of radar signals with the overall recognition results. The GCPS includes the CNN sensitive region of the TFI and the relationship between the correctly identified areas and the sensitive areas. Converting this relationship into numbers can reveal CNN's performance identifying different types of radar signals and reflect the recognition rate's reliability.

Fig. 6(c) and Fig. 7 show that the increase in SNR can significantly reduce the number of recognition errors made by the two CNNs. From Fig. 6(c), when SNR = -4dB, the recognition rate of GoogLeNet exceeds 95%. However, Fig. 7(b) shows that only 90 radar signals with SFM are correctly recognized. It has a lower recognition rate than other signals.

Fig. 6(a) and Fig. 6(b) show that the GCPS-IB is close to 0.6, and the GCPS-EB is close to 0.06. Compared with other types of signals, GoogLeNet has a low recognition effect on SFM and causes a decrease in the GCPS-IB and an increase in the GCPS-EB. This point is in the second quadrant of Fig. 6(d) and is close to the third quadrant. Although the recognition rate of GoogLeNet reached 95%, the reliability of the recognition rate is not the highest, and the network still needs to be optimized.

IV. CONCLUSION

In this letter, we proposed the GCPS, which consists of the GCPS-IB and GCPS-EB, to evaluate the CNN performance. We used five different modulation types of radar signals and two widely used CNNs to verify the performance of GCPS. The results show that GCPS-IB is not quickly saturated, and its acquisition does not require many test sets. GCPS-EB can reveal the reason why SNR affects recognition performance, not simply demonstrating the relationship between SNR and CNN performance. Therefore, compared with the recognition rate, GCPS can evaluate CNN performance more comprehensively and accurately.

REFERENCES

[1] M. Gupta, G. Hareesh, and A. K. Mahla, "Electronic warfare: Issues and challenges for emitter classification," *Defence Science Journal*, vol. 61, no. 3, p. 228, 2011.

[2] G. Zhang, L. Hu, and W. Jin, "Complexity feature extraction of radar emitter signals," in *Asia-Pacific Conference on Environmental Electromagnetics, 2003. CEEM 2003. Proceedings.* IEEE, 2003, pp. 495–498.

[3] S.-H. Kong, M. Kim, L. M. Hoang, and E. Kim, "Automatic lpi radar waveform recognition using cnn," *Ieee Access*, vol. 6, pp. 4207–4219, 2018.

[4] I. Hadji and R. P. Wildes, "What do we understand about convolutional networks?" *arXiv preprint arXiv:1803.08834*, 2018.

[5] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[6] Q.-s. Zhang and S.-C. Zhu, "Visual interpretability for deep learning: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 27–39, 2018.

[7] S. Stankovic, I. Djurovic, and V. Vukovic, "System architecture for space-frequency image analysis," *Electronics Letters*, vol. 34, no. 23, pp. 2224–2225, 1998.

[8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] N. R. Mann, N. D. Singpurwalla, and R. E. Schafer, "Methods for statistical analysis of reliability and life data," 1974.