Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33151

Recognition Robustness and Stability for Multiple Types of Neural Networks

Zongcheng Qiu

Faculty of Information Science and Engineering, Ocean University of China, Qingdao, China
Siowoqwq@outlook.com

Abstract. Neural networks have been widely used in image recognition and other practical applications, and have shown excellent performance. However, the problem of insufficient robustness and stability exists in neural networks. Under the disturbance of noise and counter samples, the performance of the model decreases significantly, which affects the safety and reliability of the actual scene. This article focuses on Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). The application scenarios and core principles of the robust optimization method are combined. Through the comparative analysis of several research cases, this paper obtains the application effect and practical advantages of various robust optimization methods. In this paper, several kinds of robust optimization methods are analyzed combined with defensive distillation, and their limitations in practical application are found. The improvement directions for these limitations are proposed, and the feasibility of these improvement directions is confirmed. By comparing and analyzing the performance of various robust optimization methods in several practical application cases, this paper concludes that these methods have the advantages of resisting noise interference and sample attack. However, there are also some limitations, such as strong specificity of adaptation scenarios, weak cross-domain migration ability, single defense dimension of optimization methods, weak defense board, or inherent limitations of general-purpose defense methods. This paper puts forward some improvement directions, such as using a general regularization method, using an adaptive weighted loss function, using confrontation training for collaborative assistance, and finally demonstrates and analyzes the feasibility of these improvement methods.

Keywords: Convolutional neural networks, recurrent neural networks, long short-term memory networks, defensive distillation, robustness.

1. Introduction

The robustness of a neural network is one of the important indices of model training. The pursuit of neural network robustness is the basic requirement for the model to meet the engineering application. By improving the robustness of the neural network, it can ensure the stability judgment of the model in a complex environment, and promote the stable application of the model in a wider range of scenarios. Robust optimization methods are widely used in convolutional neural networks

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33151

(CNNs), cyclic neural networks, long-term and short-term memory neural networks, and many other neural networks.

Taking the dual-channel CNN as an example, the image enhancement method can not only improve the image quality but also enhance the robustness of the CNN. WANG Shulin and others built an underwater image enhancement network based on this method, and accurately extracted the details of the image by combining the dense connection and efficient channel attention mechanism. [1] The results show that the underwater image quality evaluation index (UIQM) of the network on the UIEB dataset is 3.0056, and the entropy is 7.6547 [1]. The above research shows that the dual channel image enhancement method can not only effectively restore the details and colors of the underwater image, but also improve the adaptability of the model to the complex environment, which provides a guarantee for the application of the model in the actual underwater scene.

However, the existing studies lack information on the adaptation scenario, defense dimension, and inherent defects of the robustness optimization method. Many robust optimization methods in research only cover limited application fields, and different algorithms are used to improve the robustness of neural networks.

In this paper, the robustness optimization methods used in the practical application of neural networks, such as CNN, RNN, and LSTM, are studied and analyzed. Combined with the general robustness improvement methods such as defensive distillation, the advantages and limitations of these practice methods are compared and analyzed, and the improvement methods are proposed. The purpose is to improve the comprehensive understanding of robust optimization methods and their rational use in practical applications.

2. Case studies

2.1. Application of a CNN in image recognition and its robustness optimization method

CNN is a feedforward neural network. The core design idea of the network is that each neuron perceives only the local image through the local receptive field, and then combines the local information to obtain the global information through weighted summation. Finally, the neural network relies on the softmax layer to obtain the required probability data distribution. The principle of the network is to reduce the size of the image by down-sampling, which makes use of the translation invariance and scaling invariance of the image mode. The network calculates the convolution through the convolution layer composed of an input layer, a receptive field, and an output layer to realize the feature mapping, and then divides the obtained feature mapping into multiple regions by pooling, and extracts the average or maximum value of all neurons in each region.

Researchers such as Wang Shulin proposed a dual-channel underwater image enhancement CNN, which can achieve image enhancement for blurry, color cast, and low contrast underwater raw images, providing diverse samples for underwater target recognition, feature extraction, and matching models [1]. The experiment shows that compared with other algorithms, this network achieves a UIQM of 3.0056 and an image information entropy of 7.6547 on the UIEB dataset, both ranking first, proving that it can still preserve richer image details and texture information in noisy environments. On the non-training dataset of EUVP, the UIQM reached 3.1386, PSNR reached 24.3499, and SSIM reached 0.8323, demonstrating strong generalization ability and robustness in unknown noise scenarios [1]. The underwater image enhanced by this network obtained 305 matching points in GMS feature point matching, significantly higher than other comparison algorithms, effectively improving the robustness and accuracy of underwater target feature

DOI: 10.54254/2755-2721/2026.GU33151

extraction and matching, and providing more reliable input for underwater target recognition in noisy environments [1].

Researchers such as Huang Ying used a CNN-based tea tree sprout recognition model to address interference issues such as lighting changes, occlusion, and complex backgrounds in natural tea gardens. They enhanced the edge and texture features of sprouts through the edge information enhancement module EIE, and focused on key target features through channel priority convolutional attention CPCA to suppress noise. They also used the content-aware feature recombination CARAFE to ensure recognition accuracy under noise interference, such as occlusion and small targets [2]. The experimental results show that the image preprocessing enhancement strategy achieves a model accuracy of 92.1%, a recall rate of 87.4%, and an average accuracy mean mAP@0.5 Reaching 94.8%, which is 2.9, 2.6, and 2.7 percentage points higher than the original YOLOv8n, fully proves that the combination of CNNs and image enhancement can significantly improve the accuracy and robustness of object recognition [2].

In summary, in tasks such as image recognition, image enhancement can be used during model training to provide higher-quality inputs for object recognition, detection, and tracking, improving the generalization ability and robustness of neural networks. The principle of image enhancement is to highlight or suppress certain information in the image according to specific needs, in order to improve visual effects or make the image more conducive to subsequent analysis. The use of image enhancement methods may not create new data, but it can change the presentation of images. [3] Essentially, it is to enhance the diversity of effective samples, so that the model is not mechanically memorizing training samples, but learning to remember the universal features of the samples, such as overall features and contours, thereby reducing overfitting and improving generalization ability [3]. And these technologies simulate real-world interference, enabling the model to recognize various imperfect images, achieve noise-resistant recognition, and stabilize against adversarial samples.

2.2. Application of recurrent neural networks in industrial equipment state detection and its robust optimization method

RNN is a feedback network. When data is calculated, it can not only process circularly, but also use the previous output data. The principle of the network is to process arbitrary-length sequences by using neurons with feedback. The nodes between the hidden layers in the network are connected, and the input of the hidden layer includes the input layer and the output of the hidden layer at the previous time. There is a weight value for calculation from the previous neuron to the next neuron, as well as the hidden layer and the previously input hidden layer. Data transfer between neurons requires the use of activation functions to convert linear data into nonlinear data, so as to expand the coverage of data.

Medsker and other researchers have realized a one-to-one correspondence between the oscillation mode of the model and the dynamic time series data based on the cyclic neural network of chaotic oscillation diffraction, which can better process the dynamic time series data, such as voice and video [4]. Aiming at the problems of vulnerability to noise interference, large dynamic playback, and insufficient model robustness in time series signal processing, the chaotic network's oscillatory diffraction characteristics are used to stabilize the network state and feature map, effectively suppress noise interference, and stabilize the time series feature memory and attractor structure [4]. In the prediction and dynamic identification of noisy time series, the robustness of the model is improved by more than 30%, and the error response rate is reduced by 45%, which significantly improves the stability and generalization ability of RNN in the face of interference [4].

DOI: 10.54254/2755-2721/2026.GU33151

When detecting the status of industrial equipment, it is necessary to process long-term time-series data with noise and non-stationary characteristics. The traditional RNN will disappear due to the cascade effect during back propagation, and it is unable to capture the long-term impact caused by early failures. The NARX network can shorten the propagation path of the gradient and alleviate information forgetting by displaying the introduction of historical input/output. On this basis, GE Jiahao and other researchers proposed the NARX network cyclic neural network based on the combination of forward difference and improved wavelet de-noising. Aiming at the problems of noise interference, non-stationary fluctuation, and insufficient robustness of such noisy non-stationary chaotic sequences, they improved the data quality through differential stabilization and wavelet packet de-noising, and combined with closed-loop NARX to enhance the network memory and anti-interference ability [5]. The experimental results show that the RMSE of the model is as low as 2.36×10^{-2}, the training speed is faster, and the accuracy is higher, which significantly improves the prediction robustness and online application ability of RNN in a noisy environment [5].

To sum up, the robustness of neural networks can be improved through the optimization of neuron structure and dynamic mechanisms in the processing of time series data and state detection tasks. The NARX network can alleviate the early information forgetting problem caused by the disappearance of the gradient in traditional RNNs. The mixed segment oscillating diffraction can enhance the adaptability of the network to noise and disturbance by virtue of the dynamic diffraction characteristics of the chaotic state. Both methods improve the recognition stability of the model for non-stationary, noisy, and disturbed data by optimizing the time-series feature expression and information transfer mode, and improve the generalization ability of RNN and robustness in complex environments.

2.3. Application of long and short memory neural network (LSTM) in speech recognition and its robustness optimization method

LSTM is a neural network derived from the forgetfulness and selectivity of the brain's memory function. Based on the cyclic neural network, the network principle uses the gate mechanism, has a judgment and decision-making mechanism for information transmission, and can achieve selective memory and forgetting, rather than processing all data. This mechanism is composed of multiple gates according to the design order, including the input gate to judge whether the input information is meaningful, the forgetting gate to judge whether the historical information is important, and the output gate to judge whether the current output information needs to be updated. It can completely solve the forgetting problem of RNN, and is good at processing some dynamic sequence data, including number recognition, character recognition, physical signal processing, and can effectively integrate.

In speech recognition, Alex graves and other researchers based on two-way LSTM architecture and deep stacking design (DBLSTM), aiming at the problem that the traditional LSTM can only use historical information and cannot use future information, which leads to a significant decline in robustness in the face of voice differences, environmental noise and channel distortion, used both historical and future information through the two-way time series structure, and introduced the weighted noise regularization enhancement model during training [6]. Experimental results show that the phoneme error rate (PER) and frame error rate (FER) of the model are reduced to 17.99% and 27.88%, respectively, on the TIMIT data set [6]. In the WSJ speech recognition task, the word error rate (WER) is as low as 11.7%, and its performance is better than that of the conventional LSTM model, which fully proves that the deep bidirectional structure combined with weight

DOI: 10.54254/2755-2721/2026.GU33151

regularization can significantly improve the temporal modeling accuracy and robustness of LSTM [6].

Zhang Yueyao, based on the bidirectional LSTM (ISSA - BiLSTM) optimized by the improved sparrow algorithm, aimed at the problems of time series data in network security, such as too much noise, large fluctuation and poor robustness, captured the long-term dependence by the bidirectional structure, suppressed the over fitting by dropout, and used the improved sparrow algorithm to optimize the parameters to avoid local optimization [7]. The experimental results show that the MSE of the model is as low as 0.000885 and the R^2 is up to 0.986332. The prediction accuracy and robustness are significantly better than those of the traditional LSTM, and the robustness of the long-term and short-term memory neural network in noisy time-series tasks is enhanced [7].

To sum up, the two-way LSTM structure makes up for the insufficient use of one-way network context by simultaneously using historical and future context information. By purifying the features layer by layer, the deep stacked bidirectional LSTM can learn the core invariant features from the original sequence data with noise, distortion, and diversity, which significantly improves the stability and generalization ability of the model in noise, distortion, and difference scenarios, and enhances the robustness of the LSTM.

2.4. Effect of defensive distillation in the face of counter samples and improvement of its robustness optimization performance

Defensive distillation is a general technology for neural network robustness optimization. Its core principle is to improve the robustness of the neural network by using the training process of double models of teachers and students. Its core design logic is to let the student neural network fit the soft label of the teacher neural network, avoid the overfitting of the model to the training data, and then eliminate the visual blind spot caused by the high nonlinearity of the neural network.

Bintao Wang proposed a model based on the dual mechanism defense method using defensive distillation and fractional information protection. Aiming at the problems of weak generalization ability and insufficient robustness when the deep learning visual model is a single defense, gradient information concealment is realized through defensive distillation [8]. The experimental results show that the model classification accuracy of this method is 84.26% when it is used in white box attacks, such as BIM and MIM, and black box attacks, such as ZOO and AUTOZOOM, on the ImageNet dataset. The classification accuracy of clean samples on cifar10 data set is 86.37%, which fully proves that defensive distillation can effectively improve the robustness of the neural network and the adaptability of the model to white box and black box attacks [8].

Zhinan Ding proposed, based on the dual channel defensive distillation model, aiming at the problems of hidden tampering attack, poor robustness, and bulky model in abnormal power consumption detection, he improved the anti-interference ability of the model by learning the characteristics of normal and malicious samples respectively by double teachers, combining entropy balance and knowledge distillation [9].Experimental data show that the accuracy rate is 94% and AUC is 0.92 under the traditional attack, and 91% and AUC is 0.90 under the covert attack, which makes the student model significantly lightweight and effectively improves the robustness and generalization ability of the anomaly detection model [9].

To sum up, defensive distillation can form effective protection against early attack algorithms based on gradient disappearance and objective function design defects, and enable the distilled network to maintain the original classification accuracy and have certain defense capability without affecting the normal task performance. It is universal and can be applied to any feedforward neural

DOI: 10.54254/2755-2721/2026.GU33151

network. It only needs one retraining step without modifying the network architecture. It is applicable to a wide range of scenarios.

3. Discussion

3.1. Challenges and limitations

Most of the existing neural network robustness optimization methods have strong scene specificity and weak cross-domain migration ability. Most methods can only adapt to specific networks or application scenarios, such as image enhancement focusing on CNN's visual tasks, and bidirectional stacked LSTM focusing on speech recognition. These optimization methods cannot improve the robustness across networks and domains.

At present, the existing robustness optimization methods have a single defense dimension and weak defense. The defense targets of most methods are obviously biased, such as chaotic oscillatory diffraction and NARX networks, which are biased to resist the improvement of noise performance, and can not effectively deal with adversarial examples. Although defensive distillation can deal with the adversarial example attack, it can not effectively deal with the time series data with high noise.

The existing universal defense methods have inherent limitations. Defensive distillation can only reduce the success rate of traditional counter-sample attacks, but it can not fundamentally solve the core cause of the local linearization of neural networks. In the face of new high-intensity counter sample attacks, the defense ability is weak, and the security of practical applications cannot be guaranteed [10].

3.2. Direction of optimization

The general regularization method can build a portable robustness improvement framework independent of network architecture and application field, and adapt to a variety of neural networks, which can be applied in many fields. The adaptive regularization and Gaussian mixture regularization methods proposed by Yaqing Guo can simultaneously adapt to CNN face recognition models, LSTM time series prediction models, and other network structures [11]. On the noisy regression data set, the MSE of the model is reduced by more than 40% compared with the traditional LASSO. In face recognition tasks with extreme illumination noise, the recognition rate increased from 47.9% to 91.3% [11]. In the scenario of high-dimensional time series data, the accuracy of feature selection is improved by 35%. It is fully proven that the general regularization method can improve the robustness of the model on multi-class neural networks [11].

Through multi-target cooperative training, the recognition ability of the model against noise and the ability to defend against the sample are improved. The adaptive weighted loss function is used in training, so that the model can not only improve the stability under noise interference, but also keep the boundary smooth against disturbance. Xiaoyun Ren and other researchers proposed an adaptive weighted loss function (AWL) based on dynamic learning of noise labels, which can realize stable learning in the early stage of training and adaptive optimization of difficult samples in the late stage through dynamic weight scheduling [12]. On the CIFAR - 10 dataset with a high noise rate η=0.8, the classification accuracy of the model is improved from 35.2% of the traditional cross entropy to 73.2% [12]. On the CIFAR - 100 high noise data set, the accuracy increased from 19.3% to 53.1%, significantly better than the static loss function [12]. Moreover, AWL effectively reduces the sensitivity of the model to small adversarial disturbance by dynamically smoothing the gradient and suppressing the error gradient propagation, making up for the single defect of the traditional method

DOI: 10.54254/2755-2721/2026.GU33151

that only resists noise or only defends against adversarial examples, and realizing the dual improvement of the robustness of high noise environments and adversarial examples [12].

In order to make up for the limitations of defensive distillation, it uses confrontation training to carry out targeted training on the emerging new high-intensity confrontation samples. The robust visual question answering method (DDVQA) proposed by Desen Yuan, based on defensive self-distillation and the same confrontation training, can well demonstrate the feasibility of this method. While using defensive distillation to smooth the model gradient and reduce the confrontation sensitivity, it introduces confrontation training [13]. Actively add multimodal countermeasure samples in the training to ensure the output consistency when facing the original samples and countermeasure samples, and let the model learn the core invariant features [13]. The experimental results show that when the VQA-CP v2 data set is faced with multimodal countermeasures such as FGSM and PGD, the overall accuracy of the model is improved from 17.64% to 48.61%, and the accuracy of other types of problems is improved from 2.39% to 25.59% [13]. The accuracy rate of the model in the non-attack scenario is maintained at 51.89%, which is significantly better than that using only the defensive distillation method, which fully proves that when the defensive distillation is used in conjunction with the countermeasure sample, the robustness of the model against high-intensity countermeasure samples can be greatly improved [13].

4. Conclusion

This paper focuses on the robustness and stability of CNN, RNN, LSTM, and other neural networks. It sorts out the application logic and core principles of these neural networks in the fields of image recognition, industrial equipment detection, speech recognition, and so on. It analyzes the advantages and limitations of the specific robustness optimization methods of various networks and the general optimization methods of defensive distillation, and puts forward the optimization scheme.

The research shows that the lack of robustness of neural networks is due to the inherent defects of the architecture and external disturbance. For example, the feedforward structure of CNN is sensitive to dynamic disturbance, the RNN chain structure is prone to gradient disappearance, and the LSTM one-way shallow design is difficult to distinguish between core features and redundant noise, which are further exacerbated by external factors such as natural noise and countermeasure samples.

Although the existing robust optimization methods have their own effects, they still have obvious limitations: strong scene specificity, weak cross-domain migration ability, single defense dimension, weak defense weaknesses, and inherent limitations of universal defense methods. In the future, it can focus on and use the regularization method to improve the model robustness of multi-class neural networks, use multi-objective cooperative training and adaptive weighted loss function to achieve the dual improvement of the stability of the model against noise disturbance and the robustness against the attack of countermeasure samples, and use the defensive distillation and countermeasure samples to improve the robustness of the model against high-intensity countermeasure samples.

References

[1] Wang, S., Yang, J., Lu, C., & Liu, L. (2023). An enhanced CNN for underwater images based on dual channels. Marine Engineering, 41(6), 158–170.

[2] Huang, Y., Li, L., Yang, X., & Yan, J. (2025). Research on identification of different grade tea buds based on CNN. China Agricultural Science and Technology Guide, 1–11.

DOI: 10.54254/2755-2721/2026.GU33151

[3] Maini, R., & Aggarwal, H. (2010). A comprehensive review of image enhancement techniques. arXiv preprint arXiv: 1003.4053.
[4] Medsker, L. R., & Jain, L. C. (Eds.). (2001). Recurrent neural networks: Design and applications. CRC Press.
[5] Ge, J., Xiang, J., & Li, D. (2024). Online prediction of non-stationary chaotic time series with noise based on combinational NARX neural network. Acta Aeronautica et Astronautica Sinica, 45(21), 301–314.
[6] Graves, A., Jaitly, N., & Mohamed, A. R. (2013, December). Hybrid speech recognition with deep bidirectional LSTM. In 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (pp. 273–278). IEEE.
[7] Zhang, Y. (2023). Research on network security situation prediction model based on BiLSTM (Master's thesis, North China University of Technology).
[8] Wang, B. (2024). Research on adversarial attack and defense for deep learning vision model (Master's thesis, University of Electronic Science and Technology of China).
[9] Ding, Z. (2024). Research on covert data tampering attack and defense under abnormal electricity detection (Master's thesis, Taiyuan University of Science and Technology).
[10] Carlini, N., & Wagner, D. (2017, May). Towards evaluating the robustness of neural networks. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 39-57). IEEE.
[11] Guo, Y. (2023). Regularized robust regression modeling methods based on analyzing characters of noise (Doctoral dissertation, Shanxi University).
[12] Ren, X., & Tao, Q. (2025). Adaptive weighted loss function based on noise label dynamic learning. Electronic Production, 33(23), 28–33.
[13] Yuan, D. (2023). Research on methods of visual question answering based on adaptive multimodal feature fusion (Master's thesis, University of Electronic Science and Technology of China).