Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/100/2025.18828

AIGC Detection Model Based on Capsule Networks

Jiarui Zhu[1,a,*]

[1]*Institute of Internet of Things Engineering, Wuxi University, 333 Xishan Avenue, Wuxi, Jiangsu, China*
a. ray1214@foxmail.com
**corresponding author*

Abstract: With the advancement of technology, artificial intelligence-generated content (AIGC) has facilitated people's lives while also giving rise to numerous issues. Traditional AIGC detection methods have suffered from low accuracy and other problems, rendering them ineffective in detecting AI-generated images. Meanwhile, models trained on large datasets are constrained by the dataset size. Recent research has demonstrated that although training-free models are efficacious, their generalization ability poses a problem. In this paper, we propose a model based on capsule neural networks. The capsule network model acquires the spatial features of fake images and outputs image classification results via softmax classifier. We trained and evaluated the proposed AIGC image detection model using the publicly available MINIST dataset. The experimental results indicate that the capsule network-based model surpasses many traditional AIGC image detection models.

Keywords: Fake images detection, Capsule neural networks, Image classification.

1. Introduction

With the advancement of technology, AI-generated content (AIGC) is extensively employed in daily production and life, conferring convenience upon people. Nevertheless, since the advent of Generative Adversarial Networks (GAN) and various image forgery technologies capable of generating highly realistic human faces, landscapes, objects, and paintings, this technology has also engendered numerous problems in human society [1][2][3]. In 2022, an artwork generated by artificial intelligence, titled "Space Opera House (Théâtre D'opéra Spatial)", claimed an art award at the annual art competition held at the Colorado State Fair [4], the acknowledgement of AI-generated artworks has kindled discussions within the art circle, for instance, the problem of copyright infringement emerging from the utilization of unauthorized images for training [5]. Moreover, AI is capable of generating counterfeit images of specific individuals to disseminate false information, and people frequently encounter difficulties in differentiating between images produced by AI and those captured by human photographers, thereby involuntarily accepting false information [6]. To make matters worse, the outcomes of AI image generation might intensify stereotypes and biases, such as racial discrimination and gender discrimination [7][8].

Therefore, the identification of AI-generated images is of crucial significance. Nevertheless, with the development of AIGC detection models, the following three issues must not be disregarded:

· The accuracy of AIGC detection models is subpar, and it poses a challenge to define precise boundaries when the picture features a rich scene.

DOI: 10.54254/2755-2721/100/2025.18828

· Related models entail a considerable amount of time for training and consume a significant quantity of human and material resources. There is a scarcity of models that can accurately identify AIGC with a small dataset.

· The generalization capacity of the model is feeble, and it constitutes a challenge to accurately detect the content generated by different models using a single model.

This paper presents a Capsule networks-based AIGC detection model. The following innovative points are put forward:

· The model is lightweight enough to operate in a wide range of environments.

· The model can be trained with a small dataset to attain high precision and be capable of recognizing complex images.

· The model possesses generalization ability in detecting images generated by different models.

The subsequent content of this paper is structured as follows: In section 2, an extensive overview of the fake image detection method is provided. After that, section 3 systematically describes the architecture of the proposed system. Later, the experimental results are shown in section 4. At last, section 5 draws relevant conclusions.

2. Related Work

The existing image detection methods for AIGC primarily depend on deep learning, which frequently demands the deployment of large-scale training. Prior to this, people mainly detected counterfeit images by inspecting the metadata of the original images. Additionally, training-free models have increasingly emerged as a popular research focus recently. Hence, this paper categorizes the detection methods into three types: traditional AIGC image detection, AIGC detection with large-scale training deployment, and training-free AIGC image detection.

2.1. Traditional AIGC Image Detection

Early studies frequently employed metadata and manually processed image features such as translation, rotation, blurring, stretching, and shrinking for detection. Nevertheless, GAN and other technologies can directly generate new images and metadata. GAN comprises a generator and a discriminator, where the generator is accountable for creating images and the discriminator is responsible for ascertaining whether an image is genuine or counterfeit. These components are learned through adversarial training. Hence, metadata-based detection approaches are ineffective for GAN-generated counterfeit images. Yu et al. [9] contend that due to factors like training data, network structure, loss function, and parameter settings, each GAN model has a distinctive fingerprint. Inspired by Yu, numerous researchers have devised various methods to explore universal fingerprints across different generation models. Zhang et al. [10] demonstrated that there are unique artifacts in generated images. Frank et al. [11] trained detectors in the frequency domain to detect them. Liu et al. [12] noticed the texture disparities between real faces and generated faces and utilized Gram-net matrices in the network to capture crucial and long-distance texture information. These methods exhibit good performance in detecting GAN models; however, when applied to images generated by diffusion models (DM) after the advent of diffusion models, their generalization ability is weak. These spatial and frequency domain methods are not well adaptable to unseen generation models, resulting in suboptimal performance across datasets in experiments.

2.2. Large-scale Training for AIGC Detection

A considerable number of large-scale training-based approaches have overcome certain shortcomings of traditional AIGC detection methods. Through the deployment of large-scale training from generated and real images, the automatic identification of discriminative features in generated images

DOI: 10.54254/2755-2721/100/2025.18828

can be accomplished. In Wang's method [13], using 720k real images and images generated by ProGAN [14] to train ResNet50 [15] can be extended to detect some image generation models. Based on this, Gragnaniello et al. [16] enhanced the detection performance by reducing two downsampling layers with an improved ResNet50 network. Woo et al. [17] put forward a dual attention false detection fine-tuning network (DA-FDFtNet), a neural network fine-tuning architecture, which manifested significant enhancements for FaceForensics++ and GAN-generated images. Meanwhile, Xi et al. [18] proposed DALL·E and DreamStudio, a novel cross-attention enhanced dual-stream network specifically designed for AI image generation and normal photo detection, consisting of two streams: residual stream and content stream, which demonstrated remarkable superiority in detecting image generation techniques from text. Nevertheless, large-scale training is more prone to the influence of the generated images in the training set, and there are considerable differences when training the same model for different datasets.

2.3. Training-free

Owing to the manifest drawbacks of the generative models trained on a large-scale basis, numerous scholars have shifted their focus to training-free models in an attempt to address the issue arising from the dataset. AEROBLADE [19] merely detects the generated images based on the reconstruction error of the images via the autoencoder. Nevertheless, it is only applicable to the images generated by the LDM employing similar autoencoders, and its generalization capability remains a challenge. Recently, Li et al. [20] proposed a training-free attribution approach that utilizes advanced prompt reconstruction and feature extraction tools to yield attribution performance comparable to that of the state-of-the-art methods. It is worthy of mention that Li et al.'s method does not impose a limit on the number of models to be tracked, thus it can adapt to the AIGC attribution problem in open environments to a certain degree, resolving the generalization problem to some extent. Meanwhile, Tan [21] et al. put forward the (DIO) framework, a straightforward and effective method with superior performance to the existing state-of-the-art methods, and extensive experiments on 33 generative models verified the generalization ability of the proposed DIO in extracting general artifact representations. He et al. [22] also proposed a novel method, RIGID, which identifies whether an image is AI-generated by comparing the similarity of the representations of the original and noisy distorted counterparts. Furthermore, RIGID demonstrates a strong generalization ability in terms of different image generation methods and robustness against image damage. It is notable that although the aforementioned models possess a certain degree of generalization ability, the challenges of generalization ability and accuracy persist for the increasingly potential AIGC generative models.

Therefore, with the continuous development of AIGC, traditional AIGC detection models lack sufficient accuracy to recognize images generated by more advanced and theoretically supported new models. While training-based AIGC detection models can identify these models to a certain extent, they still have inadequate generalization across different models. Not only are the training effects readily influenced by the dataset, but they also demand a considerable amount of data for training, entailing a high cost. Training-free models address the issue of the dataset; however, their weak generalization and accuracy pose yet another challenge.

3. The Proposed Model

In this section, this paper puts forward the utilization of Capsule networks in place of neurons [23] for the identification of AIGC models. The details of the model are presented in Figure 1.

DOI: 10.54254/2755-2721/100/2025.18828

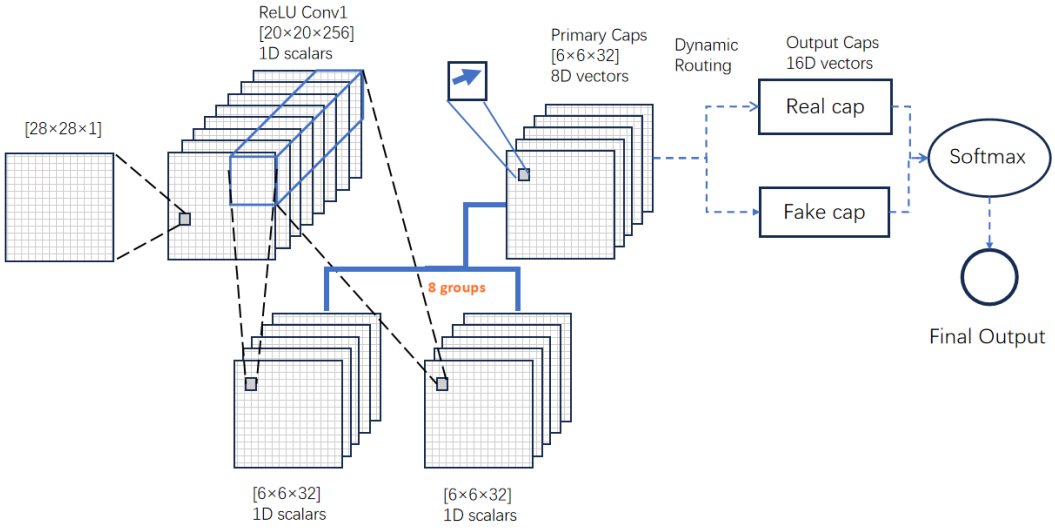

Figure 1: Capsule networks Model.

3.1. Model Details

The Capsule Networks encompass several primary capsules and two output capsules ("real" and "fake"), as depicted in Figure 1. The quantity of primary capsules is unrestricted. Experiments have indicated that a considerable number of primary capsules can enhance the network performance, but at the expense of greater computational power. After the image is inputted, the subsequent steps are executed.

1) Convolution Operation: The Capsule Networks model takes the image matrix as input, which has a size of 28×28×1 and a range of 0 - 255. Hence, we initially normalize the image matrix to confine the range to [0, 1]. In the Conv1 layer, 256 9×9 convolution kernels are employed to conduct a convolution operation with a stride of 1 on the image matrix, generating 256 feature matrices of size 20×20. Subsequently, the Primary Capsules convolution layer is utilized as the input to construct the tensor structure. We apply 8 distinct weighted Con2d operations on the 256 flow feature matrices and construct 32 9×9 convolution kernels with a stride of 2 in each Con2d to accomplish the convolution operation. Eventually, 6×6×32 8-dimensional vectors are produced, each of which is a new capsule unit composed of 8 common convolutional units. The norm of the vector represents the probability of the image belonging to a specific class. The direction of the feature vector indicates the attribute extracted by the capsule.

2) Dynamic Routing: The Output Capsules layer updates the input capsules. The processing of capsules is divided into two steps: linear combination and routing. In the case of linear combination, the output activity vector u_i of the lower capsule is multiplied by the weight matrix W_{ij} to obtain the prediction vector $\hat{u}_{j|i}$.The input of the upper capsule s_j is the weighted summation of all the prediction vectors. The specific operation will be elaborated in detail in the next section. In experience, to stabilize the training process, the dynamic routing algorithm should employ three routings (R=3), which assists the primary capsules in learning different parameters.

3) Output: For the purpose of calculating the predicted labels $\hat{y}$ we apply the softmax function, Equation 1, to each dimension of the output capsule vector, with the aim of attaining stronger polarization, instead of merely using the length of the output capsule [23]

$$\hat{y} = \frac{1}{m}\sum_i softmax\left(\begin{bmatrix}\mathbf{V}^{(1)\mathrm{T}}\\ \mathbf{V}^{(2)\mathrm{T}}\end{bmatrix}_{:,i}\right) \tag{1}$$

DOI: 10.54254/2755-2721/100/2025.18828

The final outcome is the average of all the softmax outputs: The network utilizes the vector of outputs to ascertain whether the picture is genuine or a counterfeit one generated by AI.

We merely employ the cross-entropy loss function (Equation 2) to optimize the network.

$$L = -\left(y\log(\hat{y}) + (1-y)\log(1-\hat{y})\right) \tag{2}$$

Wherein y represents the true label, $\hat{y}$ represents the predicted label, and m denotes the dimension of the output capsule $_j$.

3.2. Capsule Networks

The concept of "capsule" was introduced in Capsule Networks. Each capsule constitutes a small neural network capable of recognizing specific types of visual patterns and encoding the probability and pose parameters of its existence. Through this design, capsules can retain more spatial hierarchical information.

Capsule Networks also introduce a mechanism dubbed "dynamic routing." This mechanism enables information to be transmitted between different capsules, facilitating the network to better comprehend the internal composition structure and relative spatial relationships of objects.

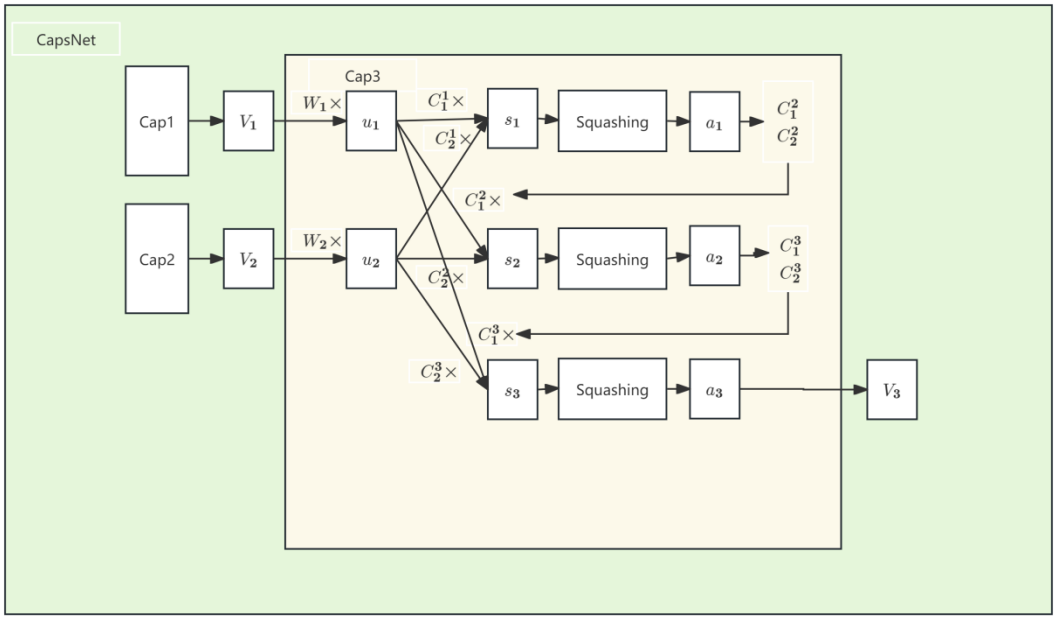

Figure 2: Example of Capsule Dynamic Routing Algorithm with R=3.

The length of the output vector of the capsule indicates the probability that the entity represented by the capsule exists in the current input. Consequently, a nonlinear "squashing" function (Equation 3) can be employed to guarantee that the length of short vectors is nearly zero, while that of long vectors is slightly less than 1. Subsequently, it is left to the discriminative learning to fully leverage this nonlinearity.

$$v_j = \frac{||s_j||^2}{1+|s_j||^2}\frac{s_j}{||s_j||} \tag{3}$$

DOI: 10.54254/2755-2721/100/2025.18828

Where v_j represents the vector output of capsule j, and s_j denotes its total input. For all capsules other than those in the first layer, the total input of capsule s_j constitutes the weighted sum of all "prediction vectors" $\hat{u}_{j|i}$ from the lower layer capsule.

For all capsules apart from those in the first layer, the total input of capsule s_j constitutes the weighted sum of all "prediction vectors" $\hat{u}_{j|i}$from the lower layer capsule.

For all capsules other than those in the first layer, the total input of capsule s_j constitutes the weighted sum of all "prediction vectors" $\hat{u}_{j|i}$ from the lower layer capsule and is generated by multiplying the output ui of the lower layer capsule by the weight matrix W_{ij} as indicated in Equations 4 and 5.

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \tag{4}$$

$$\hat{u}_{j|i} = W_{ij} u_i \tag{5}$$

Among them, $c_{ij}$is ascertained through an iterative dynamic routing process.

The total of the coupling coefficients between capsule i and all the capsules above it amounts to 1 and is determined by a "routing softmax" function, whose initial logits b_{ij} represent the logarithmic prior probabilities of capsule i coupling to capsule j. The c_{ij} computation process is presented in Equation 6.

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{6}$$

The logarithmic prior can be learned discriminately along with all other weights simultaneously. It relies on the location and type of the two capsules, but not on the current input image. Subsequently, the initial coupling coefficients are improved iteratively by measuring the consistency between the current output V_j of each capsule j in the previous layer and the predicted $\hat{u}_{j|i}$ of capsule i.

4. Experiment

4.1. Dataset

The Capsule Networks were trained using the MINST dataset [24], which encompasses 60,000 training samples and 10,000 test samples, with each image being a 28x28 pixel grayscale one. The AIGC dataset employs a fake handwritten digit dataset generated by GAN, which contains 1,000 28x28 pixel grayscale images.

4.2. Experimental environment and experimental setup

In this paper, we employ PyCharm Community Edition 2024.2.3 as the integrated development environment, establish our virtual environment through Anaconda, and operate on a Windows 64-bit operating system with Pycharm 3.9.20 and PyTorch 2.4.1 serving as the framework. The server is a Dell G16 7630 outfitted with an NVIDIA GeForce RTX 4070 Laptop GPU and a 13th Gen Intel(R) Core(TM) i9-13900HX CPU, we undertake an experiment on AIGC image classification via Capsule Networks.

4.3. Model

We adopted four additional methods for our research.

DOI: 10.54254/2755-2721/100/2025.18828

SVM [25]: It is a support vector machine, predominantly employed to address the issue of data classification within the domain of pattern recognition and is classified as a type of supervised learning algorithm.

AlexNet [26]: It is an early and relatively straightforward neural network composed of five convolutional layers and three fully connected layers.

GoogLeNet [27]: It mainly utilizes the Inception structure, featuring convolutional kernels of different sizes, signifying distinct receptive fields. The fusion of features of different scales can yield better learning outcomes. Simultaneously, since the Inception structure employs an 11x11 convolution for dimensionality reduction before convolution, the number of parameters is significantly reduced.

ResNet [28]: As the depth of the network progressively increased, performance actually commenced to decline, and certain issues such as gradient vanishing and gradient explosion also emerged. ResNet proposed a method of learning residuals instead of directly learning the input and output between network layers to address this problem.

4.4. Experimental results of Capsule Networks

As depicted in Figure 3, we trained the Capsule Networks model with the default settings and discovered that the accuracy initially rose and subsequently declined before rising again. Since the training time and accuracy were nearly identical to those at epoch 50 after epoch 80, we utilized epoch 50 for subsequent experiments to train the Capsule Networks model.

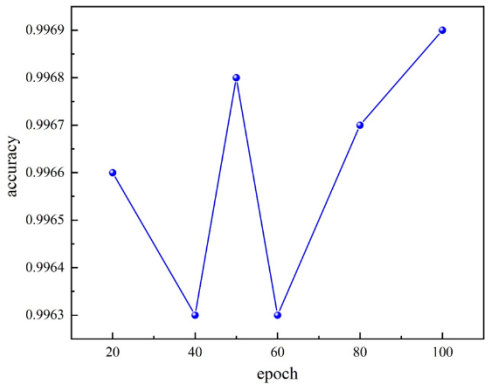

Figure 3: Accuracy versus epoch graph of Capsule Networks.

It can be observed from Figure 4 that, within the initial range of Batch size from 10 to 100, the accuracy rate demonstrates a stable fluctuating upward trend. However, it exhibits a notable downward trend after exceeding 100. Therefore, we adopt Batch size = 100 for other experiments.

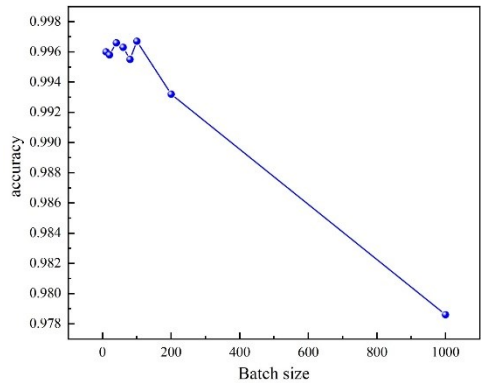

Figure 4: Accuracy versus Batch size graph of Capsule Networks.

DOI: 10.54254/2755-2721/100/2025.18828

It is evident from Figure 5 that there exists a distinct mutation point at a learning rate of 0.16, and it is noted that a good effect is achieved when the learning rate is 0.0001. In Figure 6, it is observed that when the dynamic routing number is set to 3, a favorable outcome is obtained. Hence, we ultimately configure the Capsule Networks model as Batch size = 100, epoch = 50, routings = 3, and learning rate = 0.001. From Table 1, it is apparent that the test set accuracy of the Capsule Networks fluctuates within a range of 0.001, indicating excellent stability.

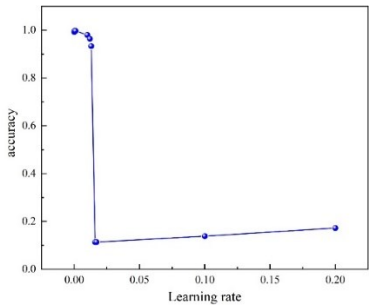

Figure 5: Accuracy versus learning rate graph of Capsule Networks.

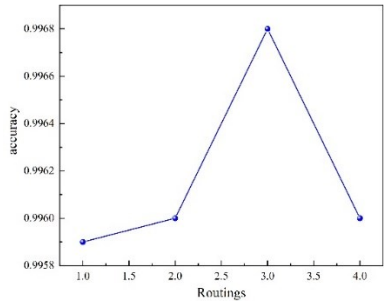

Figure 6: Accuracy versus routings graph of Capsule Networks.

Table 1: Accuracy of test set when the model of Capsule Networks is set to Batch size=100, epoch=50, routings=3, and learning rate=0.001

Training number	Test set accuracy
1	0.9963
2	0.9968
3	0.9966
4	0.9969
5	0.9966

4.5. Control group experiment

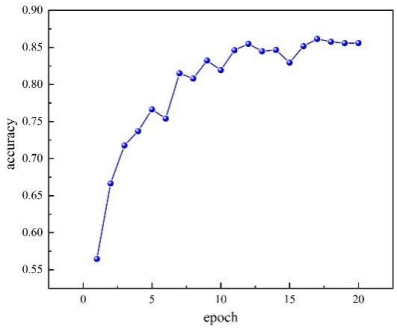

Figure 7: Line chart of AlexNet accuracy.

DOI: 10.54254/2755-2721/100/2025.18828

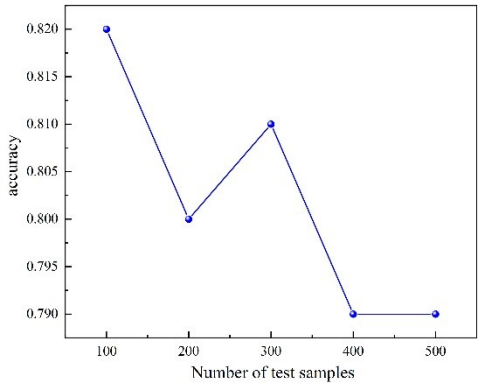

Figure 8: Line chart of SVM accuracy.

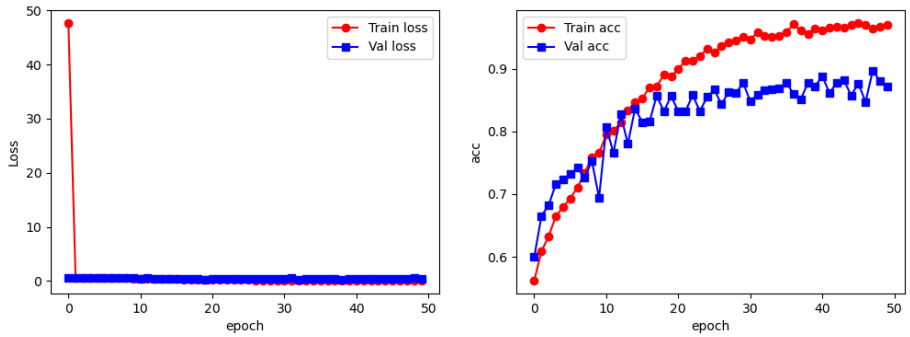

Figure 9: Line chart of accuracy and loss of GoogLeNet.

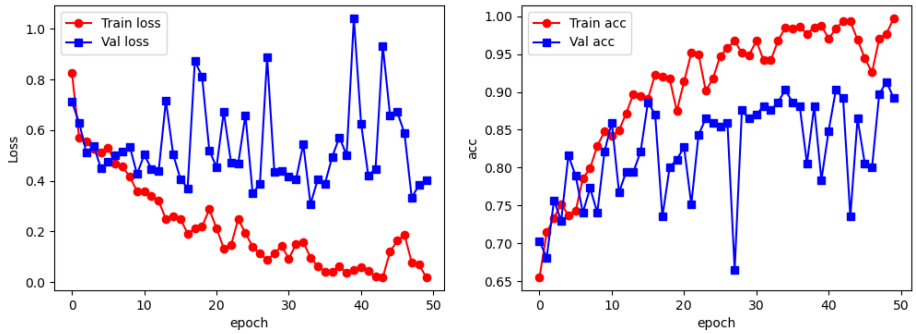

Figure 10: Line chart of Resnet model accuracy and loss.

Table 2: Comparison results between Capsule Networks and other models

Model	Test set accuracy
Capsule Networks	0.9969
SVM	0.8200
AlexNet	0.8614
GoogLeNet	0.8966
Resnet	0.9135

It is observable from Table 2 that the model accuracy of Capsule Networks is the highest. In comparison with the SVM image classification model, the accuracy of Capsule Networks is

DOI: 10.54254/2755-2721/100/2025.18828

approximately 17.5% higher; meanwhile, when compared to Alexnet, it is around 13.4% higher. It is clearly discernible from Fig. 9 and Fig. 10 that the training accuracy and test accuracy of the GoogLeNet model and the Resnet model increase progressively as the number of epochs rises. However, both of them exhibit the phenomenon of overfitting, and Resnet has a considerable degree of instability points. Nevertheless, Capsule Networks not only has no overfitting phenomenon but also its test set accuracy is higher than that of Resnet and GoogLeNet by 8.5% and 10.0% respectively. In conclusion, the AIGC detection model based on Capsule Networks has attained favorable outcomes, and its experimental results are superior to those of the other comparison models, confirming the efficacy of Capsule Networks for AIGC detection.

5. Conclusion

Based on the current research status of AIGC detection, we propose an AIGC detection method based on Capsule Networks. This method analyses how features are extracted from the input image at each level of capsules and delineates the relationship between each capsule and the entire network. It also offers a detailed analysis of how capsules operate. The experimental results indicate that the detection method based on Capsule Networks can train a high-accuracy model with a small quantity of data samples and does not encounter overfitting as do other models. By analyzing these experimental results and related data, we contend that this work will stimulate further research and development of capsule Networks, not only for AIGC detection but also for numerous other fields. This holds great significance for some fields with high requirements for explainability, such as medicine and finance.

References

[1] *Chen, C., Fu, J., Lyu, L.: A pathway towards responsible AI generated content. https://doi. org/10.48550/arXiv. 2303.01325*

[2] *Whittaker, L., Kietzmann, T.C., Kietzmann, J., Dabirian, A.: "All around me are synthetic faces": the mad world of AI-generated media. IT Prof. 22, 90–99 (2020)*

[3] *Sha, Z., Li, Z., Yu, N., Zhang, Y.: DE-FAKE: detection and attribution of fake images generated by text-to-image generation models. https://doi.org/10.48550/arXiv.2210.06998*

[4] *https://www.vice.com/en/article/bvmvqm/an-ai-generated-artwork-won-first-place-at-a-state-fair-fine-arts-competition-and-artists-are-pissed*

[5] *Guangzhou Internet Court, Civil Judgment No. (2024) Guangdong 0192-0113*

[6] *Pingliang City Kongtong District People's Court, Criminal Judgment No. 105 (2024) Gansu 0802*

[7] *https://www.theverge.com/2021/9/4/22657026/facebook-mislabeling-video-black-men-primates-algorithm*

[8] *https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/*

[9] *YU N, DAVIS L, FRITZ M. Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints[C/OL]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South). 2019. http://dx.doi.org/10.1109/iccv.2019.00765. DOI:10.1109/iccv.2019.00765.*

[10] *ZHANG X, KARAMAN S, CHANG S F. Detecting and Simulating Artifacts in GAN Fake Images[C/OL]//2019 IEEE International Workshop on Information Forensics and Security (WIFS), Delft, Netherlands. 2019. http://dx.doi.org/10.1109/wifs47025.2019.9035107. DOI:10.1109/wifs47025.2019.9035107.*

[11] *FRANK J, EISENHOFER T, SCHÖNHERR L, et al. Leveraging Frequency Analysis for Deep Fake Image Recognition[J]. arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition, 2020.*

[12] *LIU Z, QI X, TORR P H S. Global Texture Enhancement for Fake Face Detection in the Wild[C/OL]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. 2020. http://dx.doi.org/10.1109/cvpr42600.2020.00808. DOI:10.1109/cvpr42600.2020.00808.*

[13] *Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. CNN-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8695–8704, 2020.*

[14] *KARRAS T, AILA T, LAINE S, et al. Progressive Growing of GANs for Improved Quality, Stability, and Variation[J]. Learning, Learning, 2017.*

DOI: 10.54254/2755-2721/100/2025.18828

[15] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016. http://dx.doi.org/10.1109/cvpr.2016.90. DOI:10.1109/cvpr.2016.90.
[16] GRAGNANIELLO D, COZZOLINO D, MARRA F, et al. Are GAN generated images easy to detect? A critical analysis of the state-of-the-art[J]. arXiv: Computer Vision and Pattern Recognition, arXiv: Computer Vision and Pattern Recognition, 2021.
[17] BANG Y, WOO Simon S. DA-FDFtNet: Dual Attention Fake Detection Fine-tuning Network to Detect Various AI-Generated Fake Images[J].
[18] AI-Generated Image Detection using a Cross-Attention Enhanced Dual-Stream Network.pdf[J].
[19] RICKER J, LUKOVNIKOV D, FISCHER A. AEROBLADE: Training-Free Detection of Latent Diffusion Images Using Autoencoder Reconstruction Error[J]. 2024.
[20] LI M, QIAN Z, ZHANG X. Regeneration Based Training-free Attribution of Fake Images Generated by Text-to-Image Generative Models[J].
[21] TAN C, LIU P, TAO R, et al. Data-Independent Operator: A Training-Free Artifact Representation Extractor for Generalizable Deepfake Detection[J].
[22] HE Z, CHEN Y, HO T Y. RIGID: A Training-Free and Model-Agnostic Framework for Robust AI-Generated Image Detection[J].
[23] SABOUR S, FROSST N, HINTON Geoffrey E. Dynamic Routing Between Capsules[J]. Neural Information Processing Systems, Neural Information Processing Systems, 2017.
[24] https://yann.lecun.com/exdb/mnist/
[25] https://github.com/chestnut24/SVMImageClassification
[26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J/OL]. Communications of the ACM, 2017: 84-90. http://dx.doi.org/10.1145/3065386. DOI:10.1145/3065386.
[27] SZEGEDY C, WEI LIU, YANGQING JIA, et al. Going Deeper with Convolutions[C/OL]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA. 2015. http://dx.doi.org/10.1109/cvpr.2015.7298594. DOI:10.1109/cvpr.2015.7298594.
[28] [HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016. http://dx.doi.org/10.1109/cvpr.2016.90. DOI:10.1109/cvpr.2016.90.