ResCaps: an improved capsule network and its application in ultrasonic image classification of thyroid papillary carcinoma

Xiongzhi Ai¹ · Jiawei Zhuang¹ · Yonghua Wang¹ · Pin Wan¹ · Yu Fu²

Received: 18 October 2020 / Accepted: 17 March 2021 / Published online: 31 March 2021
© The Author(s) 2021

Abstract

Ultrasonic image examination is the first choice for the diagnosis of thyroid papillary carcinoma. However, there are some problems in the ultrasonic image of thyroid papillary carcinoma, such as poor definition, tissue overlap and low resolution, which make the ultrasonic image difficult to be diagnosed. Capsule network (CapsNet) can effectively address tissue overlap and other problems. This paper investigates a new network model based on capsule network, which is named as ResCaps network. ResCaps network uses residual modules and enhances the abstract expression of the model. The experimental results reveal that the characteristic classification accuracy of ResCaps3 network model for self-made data set of thyroid papillary carcinoma was 81.06%. Furthermore, Fashion-MNIST data set is also tested to show the reliability and validity of ResCaps network model. Notably, the ResCaps network model not only improves the accuracy of CapsNet significantly, but also provides an effective method for the classification of lesion characteristics of thyroid papillary carcinoma ultrasonic images.

Keywords Capsule network · Papillary carcinoma of the thyroid · ResCaps · Ultrasound images

Introduction

Thyroid cancer is the most common thyroid malignancy, which accounts for about 1% of systemic malignancy. As a type of thyroid cancer, thyroid papillary carcinoma accounts for the largest proportion of thyroid cancer [1], and can occur at any age. Recently, more and more people suffer from such diseases. Thus, it is especially important to improve the accuracy of diagnosis of papillary thyroid carcinoma. Ordinarily, the ultrasonic images of thyroid papillary carcinoma have low

resolution and poor definition. Moreover, the thyroid gland has the complex tissue structure and many interfering factors in the ultrasonic images. Thus, it is difficult to classify the features of the lesions in the ultrasonic images of thyroid cancer.

In recent years, the classification of medical images is usually achieved by convolutional neural network (C-NN) [2]. Sabour et al. used a CNN to distinguish the brain of Alzheimer's patients with normal and healthy brains [3]. The famous LeNet-5 network model was used to classify functional magnetic resonance imaging data of Alzheimer's patients from the normal control group, which has the accuracy rate of 96.85%. [4] proposed a method to improve the Faster RCNN scheme for detecting speed and to promote the efficiency of classification accuracy. This scheme used the convolution of fourth and fifth layer to take $L2$ regularization processing. Then, the two layers of realizing information sharing with multi-scale were connected as the input port. Finally, the accuracy and speed of the network model recognition effect is improved, respectively. Li et al. added a spatial constraint layer to CNN, which enabled the detector to extract the features around the cancer area [5]. In addition, the more obscure or smaller cancer areas could be detected by the detector, because the shallow and deep layers of CNN are connected. However, the detector contains a large number of

✉ Yonghua Wang
wangyonghua@gdut.edu.cn

Xiongzhi Ai
17607011056@163.com

Jiawei Zhuang
zhuangjiawei@mail2.gdut.edu.cn

Pin Wan
wanpin2@163.com

Yu Fu
yfu985@126.com

¹ School of Automation, Guangdong University of Technology, Guangzhou, China

² School of Zhongshan, University of Electronic Science and Technology of China, Zhongshan, China

parameters, which is not easy to transplant into the actual terminal.

In order to overcome the shortcoming that CNN cannot detect tissue overlap features effectively, Sarraf et al. proposed the capsule network model [6]. Obviously, the capsule network model increased the degree of correlation between the local and the whole. Moreover, this model represented the spatial characteristics of the object by the vector contained more information. The capsule network model is tested by the handwritten data set (MNIST) [7] and its accuracy is 99.64%. To improve these problems, such as large amount of capsule layer data and slow training speed, Geoffrey et al. proposed a matrix capsule network model [8]. The matrix capsule network model used Gaussian mixture model to complete the routing layer clustering process, which reduces the amount of data in the capsule layer. Moreover, this model is changed from a vector form to a matrix and a scalar form, where the activation value determines the probability of being selected between the upper and lower capsule layers. Thus, the experimental training speed is accelerated. Mobiny et al. proposed a rapid capsule network for the classification of lung cancer images pretreatment [9]. The fast capsule network used the same pixel value to equal the same routing correlation coefficient, which reduces the complexity of routing layer operation. Moreover, its network training speed was three times that of the capsule network. In addition, the deconvolution layer was used to replace the full connection layer of the capsule network in the rapid capsule network, which prevented overfitting and ultimately improved the classification accuracy. In summary, the capsule network is significantly better than the traditional convolutional neural network in terms of accuracy of overlapping object recognition and it contains rich information [10,11]. Therefore, the capsule network model is better than the traditional CNN in terms of the classification task.

With the continuous development of capsule network model in the field of medical imaging, this model has been fully applied in the accurate reconstruction of functional nuclear magnetic resonance [12]. A new model based on capsule network, named SegCap, was applied to the segmentation of low-dose CT scan images, which accuracy is significantly improved [13]. Recently, the capsule network model has been continuously optimized and applied. For example, CapsGan used capsule network as a discriminator of GAN, which obtained better visualization effect than the traditional GAN network based on convolution [14,15]. Moreover, DeepCaps network [16] fused the dynamic routing algorithm of capsules on the basis of deep convolutional network, which reduced the number of parameters of the original capsule network by 68% and achieving the world's leading level in CIFAR10 [17] and Fashion-MNIST open data sets [18]. Thus, the fusion of the discriminator network,

the capsule network and the deep convolutional network is also a development trend in the future.

Motivated by the above research methods, the ultrasonic image classification task of thyroid papillary carcinoma is proposed in this paper. The main contributions of this paper are reported as follows.

1. The pretreatment scheme based on the bilinear interpolation method is proposed to obtain a picture of a specified size, which satisfies the input of the capsule network. To obtain the ultrasonic image with the required resolution of 78×78 , the ultrasonic images of thyroid papillary carcinoma are processed by the proposed pretreatment scheme before inputting the network model.
2. The ResCaps network module based on the residual module and CapsNet network module is designed, which strengthens the abstract expression of the capsule network and improved the accuracy of classification tasks.
3. The ResCaps network model not only realizes the classification task a small amount of data, but also improves the classification performance of ultrasonic images of thyroid papillary carcinoma when the amount of data set is small.

Introduction to capsule network

The model structure of the capsule network is shown in (a) from Fig. 1. The capsule network is consisted of the ultrasonic image input layer, convolutional neural network layer, PrimaryCaps initial capsule layer, DigitCaps and the final output layer, which are introduced as follows:

Input layer: it is used to input the ultrasonic image of pretreated papillary carcinoma of thyroid.

Convolution layer (CONV1): the main function of conventional convolution operation is that a local feature detection of the image is performed. Low-level features are extracted by convolution operations, which provides a more applicable high-level example of the capsule layer.

PrimaryCaps initial capsule layer: the primarily low-level features can be stored as a vector. In the initial capsule layer, the spatial dimension is transformed, and the expression of the vector is obtained to prepare for the input of the next layer.

Digital capsule layer: this layer is connected with the initial capsule layer by the form of the vector to vector. Moreover, the output of this layer is calculated using a dynamic routing algorithm.

Output layer: the length of the output vector means the probability of occurrence of its represented content. Thus, the output result of classification is the $L2$ norm of each capsule vector.

Fig. 1 Different model structures

Each capsule unit of the capsule layer is represented by a vector, where a capsule unit contains the attitude parameters of the object and the probability of belonging to this class. In the traditional convolutional network, the pooling layer is a shortcoming. In the process of dimensionality reduction, the layer loses some important feature information, which leads to reducing identification accuracy. The capsule network is used in the network dynamic routing method, which is an effective way to solve the above problem [7]. For dynamic routing method, the correlation of fluctuation capsules layer can be updated by judging the weight value. Interestingly, some trade-offs between DigitCaps layer capsule unit and the proportion of the learning process are required. If the predicted results is closed to the real value, the associated

values and capsule unit of DigitCaps layer are promoted. If the predicted results and the real value are far off, the correlation value and certain capsule unit of DigitCaps layer are declined. The specific calculation process of dynamic routing is given as follows:

$$\hat{u}_{j|i} = W_{ij}u_i, \quad (1)$$

where u_i is the output of the i th capsule layer, $u_{j|i}$ means the prediction vector output of the j th layer obtained by calculating the i th capsule layer, W_{ij} is the weight value that is used for learning and back propagation:

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, \quad (2)$$

where c_{ij} is the associated value of adjacent capsule layers, b_{ij} is the probability that the i th capsule layer is selected by the j th capsule layer. The initial value of b_{ij} is set to 0, when the routing layer starts to execute:

$$s_j = \sum_i c_{ij} \hat{u}_{j|i}, \quad (3)$$

where s_j is the input vector of the j th capsule layer:

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|}, \quad (4)$$

where v_j is the output of the j th capsule layer; the output value $\frac{s_j}{\|s_j\|}$ is guaranteed to be in the interval $[0, 1]$, which does not exceed the expected range value. (4) is the remainder of the non-linear activation function:

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} v_j, \quad (5)$$

where the input $\hat{u}_{j|i}$ and output v_j are updated by the inner product between vectors. The value of b_{ij} is a parameter, which adjusts the degree of correlation between the capsule layers:

$$L_c = T_c \max(0, m^+ - \|v_c\|)^2 + \lambda(1 - T_c) \max(0, \|v_c\| - m^-)^2. \quad (6)$$

In (6), L_c is the loss function of the classified network modules. c is the category of classification. T_c is an indicator function of the classification. $T_c = 1$ when c is existence, while $T_c = 0$ when c does not exist. m^+ means the upper edge, i.e., the punishment false negative; m^- means the lower margin, i.e., false positives are penalized. λ denotes the proportional coefficient, which is used to adjust the proportion between m^+ and m^- . m^+ , m^- and λ are hyper parameters that have been set before the capsule network learning.

The convolution layer of the capsule network has a low ability to process feature information, which is not enough to provide more detailed advanced feature information for the initial capsule layer. To solve such problem, a network model with better performance is proposed in this paper, which is named ResCaps network model.

ResCaps network model

The structure of the new ResCaps network is shown in (b) from Fig. 1. Based on CapsNet network model, residual modules are fused. The size of the input image is not changed

Fig. 2 Schematic diagram of a single residual module

Fig. 3 Specific model of residual module

after the ultrasonic image passes through the residual module. The information processed by the residual module provides advanced features for the initial capsule layer better.

The principle of a single residual module is shown in Fig. 2. The residual module adds a quick connection between the input and output of the network layer, which is the identity map:

$$H(X) = F(X) + X, \quad (7)$$

where X represents the input of the residual block, $H(X)$ is the output of the residual block, and $F(X)$ is the residual:

$$F(X) = H(X) - X, \quad (8)$$

which is the deformation of (7). The residual module represents the network, which fits the residual $F(X)$ between the input and output. The residual module can deepen the network model and improve the ability of abstract expression of features in the ResCaps network module.

The specific composition of the residual module is shown in Fig. 3. The residual module is composed of modules with a convolution kernel size of 1×1 , convolution kernel size of 3×3 and channel number of 32, respectively. Here, the network

Fig. 4 Characteristics of thyroid papillary carcinoma from left to right indicate irregular shape, unclear boundary, uneven echo, calcification and normal

Fig. 5 Coordinate diagram of the bilinear interpolation method

model name is given by the number of residual modules, where one residual block is named ResCaps1, two residual blocks are named ResCaps2 and so on.

CNNCaps network model

As shown in (c) from Fig. 1, the CNNCaps network model corresponds to the ResCaps network model structure. However, their difference is the module before the PrimaryCaps. The CNN modules are consistent with the convolution parameters in the residual module. It is missing that the shortcut connection in the residual module. The CNNCaps network model can well reflect the advantages and disadvantages of the residual module and the CapsNet network model.

The CNNCaps network model is named in the same way as the ResCaps network, and the name is determined by the corresponding CNN module. For instance, the model contains a CNN module, which is named CNNCaps1. The model contains two CNN modules, which is named CNNCaps2.

Simulation analysis

Data preprocessing

There are five characteristics, such as irregular shape, unclear boundary, uneven echo, calcification and normal, which are classified as corresponding categories in simulations. Specific labeling method of thyroid papillary ultrasound image data set: 0_00001 represents the first image of this type with irregular shape, 0_00002 means the second image of this type with irregular shape, and so on. 1_00001 indicates the first image of this class with unclear boundary. 2_00001 denotes the first image of this type of uneven echo. 3_00001 represents the first image of this type of calcification. 4_00001 means the first image of the normal class. The specific classification diagram of lesion attributes is shown in Fig. 4. All kinds of ultrasonic images are made into small sample data sets.

The ultrasound image data set about thyroid papillary carcinoma contains the examination reports of 307 patients. We obtain informed consent from each patient. Moreover, each patient agrees to use their information for research, which ensures that the research process does not affect their treatment choices and does not violate their privacy. The data is collected from 2010 to 2014, in which 54 males and 253 females. A total of 4738 ultrasound images is obtained, where each person had 5–30 ultrasound images. Interestingly, among these people under investigation, 256 people are diagnosed with papillary thyroid cancer and 51 people are not. Here, 1956 images are used for training and 424 images are used for testing. Each image typically contains one to three features, and the most prominent of these features is selected as the classification result of the disease.

The image input of the ResCaps network model needs a fixed size of 78×78 resolution. Thus, we perform uniform size operation on ultrasonic images of thyroid papillary carcinoma with different sizes. These ultrasound images of thyroid papillary carcinoma are reduced by bilinear interpolation [19,20]. The coordinate diagram is used by the bilinear interpolation method, which is shown in Fig. 5. In the coordinate axis, $Q11$, $Q12$, $Q21$ and $Q22$ represent the adjacent pixel values, respectively. P is the scaled pixel value,

Fig. 6 Training accuracy of each network model

and $R1$ and $R2$ are the intermediate process values. Noted that $Q11 = (x1, y1)$, $Q12 = (x1, y2)$, $Q21 = (x2, y1)$, $Q22 = (x2, y2)$, $R1 = (x, y1)$, $R2 = (x, y2)$. The unknown function f is a pixel value. Firstly, linear interpolation is carried out in the direction of the X axis to obtain (9) and (10).

$$f(R1) = \frac{x2 - x}{x2 - x1} f(Q11) + \frac{x - x1}{x2 - x1} f(Q21). \quad (9)$$

$$f(R2) = \frac{x2 - x}{x2 - x1} f(Q12) + \frac{x - x1}{x2 - x1} f(Q22). \quad (10)$$

Then, linear interpolation is performed in the Y axis direction to obtain the (11).

$$f(P) = \frac{y2 - y}{y2 - y1} f(R1) + \frac{y - y1}{y2 - y1} f(R2). \quad (11)$$

In summary, the required pixel value of point P , $f(x, y)$, is obtained, which can be given by

$$\begin{aligned} f(P) = & \frac{f(Q11)}{(x2 - x1)(y2 - y1)} (x2 - x)(y2 - y) \\ & + \frac{f(Q21)}{(x2 - x1)(y2 - y1)} (x - x1)(y2 - y) \\ & + \frac{f(Q12)}{(x2 - x1)(y2 - y1)} (x2 - x)(y - y1) \\ & + \frac{f(Q22)}{(x2 - x1)(y2 - y1)} (x - x1)(y - y1). \end{aligned} \quad (12)$$

Simulation results and analysis

The model training uses NVIDIA RTX2080Ti with 12G memory GPU graphics card. The operating environment is Ubuntu 16.04 LTS. The accuracy of each network model is shown in Fig. 6. The horizontal axis represents the number of training steps, and the vertical axis means the accuracy of model training. The dynamic routing times of each network model are set as 3 times. Moreover, each network model batch is 24 sequential random input images. Furthermore, the final iteration time is set to 200.

Analysis of the convergence time and training speed of simulations

In Fig. 6, the simulation results are presented to conclude that the ResCaps network model increases the parameter size of the model, while the convergence time of the model is not significantly extended.

In Fig. 7, the simulation results are given to conclude that the ResCaps network model reaches a convergence in the ResCaps2, CNNCaps2, ResCaps3, and CNNCaps3 network models. Obviously, the convergence time of the above network models are small. Interestingly, all the mentioned ResCaps network model is slightly shorter than the convergence time of the CNNCaps network model. It has been proved that CNN work very well in dealing with image classification problems. However, convolution is a large number of matrix operations, and its computational cost is very large. Particularly, the capsule network can solve this problem well by the special structure, where a capsule is a nested collection of neural layers. Thus, all the mentioned ResCaps network models have shorter convergence time.

Accuracy analysis of simulation results

After running on the ultrasound image test data of thyroid papillary carcinoma, the accuracy of each model is shown in Table 1. Based on the results in Table 1, under the specific data set of papillary thyroid carcinoma, the test accuracy

Fig. 7 Training accuracy of each network model

of ResCaps3 was the highest at 81.06%. Simulation results show that the accuracy of ResCaps network model is higher than that of CapsNet.

As can be seen from the results in Table 2, CNNCaps3 has the highest test accuracy rate of 79.17% under the specific data set of papillary thyroid carcinoma. Simulations prove that the ResCaps network model has the highest accuracy than the CNNCaps network models. It is well known that CNN will lose a lot of information in the pooling layer, which reduces the spatial resolution. However, the capsule network hardly changes the resolution of the ultrasonic images of thyroid papillary carcinoma during the calculation process. Thus, the ResCaps network models have better accuracy.

Simulations based on Fashion-MNIST

In order to verify that the ResCaps network model still performs better than CapsNet network model in other data, we chose public data set Fashion-MNIST to test in simulations. The public data set of Fashion-MNIST contains 60,000 train-

ing images, 10,000 test images and corresponding labels. The data set is a visualization of human necessities, such as T-shirts, pants, and so on, which is a total of 10 classes. Each image has a resolution of 28×28 . The test accuracy of the ResCaps network model on the Fashion-MNIST data set is shown in Table 3. The dynamic routing times of each network model in the Fashion-MNIST data set is set as 3 times, and the batch of each network model was 128 randomly ordered input images, and the iteration times were 50 times.

According to the test results of Fashion-MNIST data, the ResCaps network model has the better performance than CapsNet network model. Obviously, the model structure accuracy rate of ResCaps4 is the highest at 92.26%.

Visual verification of ResCaps network model

In order to verify the convergence of ResCaps network model in the process of dynamic routing, t-SNE clustering algorithm [21] is used to reduce the dimensionality of high-dimensional data, which is displayed in three dimension coordinates.

Table 1 Accuracy of each model test of ResCaps under the thyroid papillary carcinoma data set

Model	CapsNet	ResCaps2	ResCaps3	ResCaps4	ResCaps5
Accuracy	75%	78.41%	81.06%	77.27%	77.65%

Table 2 Accuracy of each model test of CNNCaps under the thyroid papillary carcinoma data set

Model	CapsNet	CNNCaps2	CNNCaps3	CNNCaps4	CNNCaps5
Accuracy	75%	78.35%	79.17%	79.16%	77.64%

Table 3 Accuracy of each model test of ResCaps under the Fashion-MNIST

Model	CapsNet	ResCaps2	ResCaps3	ResCaps4	ResCaps5
Accuracy	90.31%	91.66%	91.97%	92.26%	92.22%

Fig. 8 Data clustering diagram of ResCap3 model training process

In Fig. 8, the dynamic clustering process of ResCaps3, the optimal model in the thyroid data set, is updated with the number of iterations. The number of iterations is increased successively from left to right and from top to bottom, which indicates the process of continuous data aggregation and the final clustering into 5 classes.

Based on all the above simulation results, it can be known that the ResCaps network model can achieve the best classification effect of capsule network when the number of ResNet modules is 3 or 4. The specific optimal number of ResNet modules varies slightly from one data set to another. Although the ResCaps network model increases the number of parameters of CapsNet model, it can be concluded from the experimental results in Figs. 6 and 7 that the training time of this model is roughly equal to that of CapNet model.

Conclusions

In this paper, the pretreatment scheme based on the bilinear interpolation method is proposed to obtain the ultrasonic images of thyroid papillary carcinoma with a specified size,

which satisfies the input of the capsule network. Furthermore, the ResCaps network model is designed, which is applicable to the classification task of thyroid papillary carcinoma. On the premise that the network model does not affect the training speed of CapsNet network model, the accuracy of the classification task is improved and reaches the optimal accuracy of 81.06%. However, the current ResCaps network model still has room for upgrading. Thus, the future research direction is how to modify the network model structure to further improve the accuracy of model classification.

Declarations

Conflict of interest All the authors have approved the manuscript for publication, and there is no conflict of interest exists.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Davies L, Welch H (2006) Increasing incidence of thyroid cancer in the United States. *JAMA* 295(18):2164–2167
2. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 29(6):82–97
3. Sabour S, Frosst N, Hinton G, Hinton GE (2017) Dynamic routing between capsules. *Adv Neural Inf Process Syst* 2017:3856–3866

4. Ke W, Wang Y, Wan P, Liu W, Li H (2017) An ultrasonic image recognition method for papillary thyroid carcinoma based on depth convolution neural network. *Neural Inf Process* 2017:82–91
5. Li H, Weng J, Shi Y, Gu W, Mao Y, Wang Y, Liu W, Zhang J (2018) An improved deep learning approach for detection of thyroid papillary cancer in approach images. *Sci Rep* 8(1):6600
6. Sarraf S, Tofighi G (2016) Deep learning-based pipeline to recognize Alzheimer's disease using fMRI data. In: *Future technologies conference*
7. Deng L (2012) The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Process Mag* 29(6):141–142
8. Geoffrey E H, Sara S, Nicholas F (2018) Matrix capsules with EM routing. In: *Conference on learning representations*
9. Mobiny A, Van Nguyen h (2018) Fast CapsNet for lung cancer screening. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Cham, 2018, pp 741–749
10. Shruthi Bhamidi SB, El-Sharkawy M (2019) Residual capsule network. In: *IEEE 10th Annual ubiquitous computing, electronics and mobile communication conference (UEMCON)*. New York, NY, USA 2019, pp 557–560
11. Shruthi Bhamidi SB, El-Sharkawy M (2020) 3-Level residual capsule network for complex datasets. In: *IEEE 11th Latin American symposium on circuits and systems (LASCAS)*. San Jose, Costa Rica 2020, pp 1–4
12. Qiao K, Zhang C, Wang L, Chen J, Zeng L, Tong L, Bi Y (2018) Accurate reconstruction of image stimuli from human fMRI based on the decoding model with capsule network architecture. *Front Neuroinform* 2018:12
13. Lalonde R, Bagci U (2018) Capsules for object segmentation. *arXiv preprint [arXiv:1804.04241](https://arxiv.org/abs/1804.04241)*
14. Jaiswal A, AbdAlmageed W, Wu Y, Premkumar N (2018) Capsule-gan: generative adversarial capsule network. In: *Proceedings of the European conference on computer vision (ECCV) workshops*
15. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan. *arXiv preprint [arXiv:1701.07875](https://arxiv.org/abs/1701.07875)*
16. Rajasegaran J, Jayasundara V, Jayasekara S, Jayasekara, H, Seneviratne S, Rodrigo R (2020), DeepCaps: going deeper with capsule networks. In: *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (2019), pp 10725–10733
17. Ho-Phuoc T (2018) CIFAR10 to compare visual recognition performance between deep neural networks and humans. In: *Computer vision and pattern recognition*
18. Xiao H, Rasul K, Vollgraf R (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint [arXiv:1708.07747](https://arxiv.org/abs/1708.07747)*
19. Wang S, Yang K (2008) Research and implementation of image scaling algorithm based on bilinear interpolation. *Autom Technol Appl* 27(7):44–45
20. Kun B, Feifei H, Cheng W (2011) An image correction method of Fisheye lens based on bilinear interpolation. In: *Fourth international conference on intelligent computation technology and automation 2011*, pp 428–431
21. Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9(9):2579–2605

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.