Proceedings of 2021 IEEE
International Conference on Mechatronics and Automation
August 8 - 11, Takamatsu, Japan

Classification of Aquatic Animals by the Spherical Amphibian Robot based on Transfer Learning

Shuxiang Guo[1,2], ShaolongWang[1]
[1]Tianjin Key Laboratory for Control Theory&Applications
In Complicated systems and Intelligent Robot Laboratory
Tianjin University of Technology
BinshuiXidao Extension
391,Tianjin,300384,China
guo@eng.kagawa-u.ac.jp;
171029492@qq.com

Jian Guo[1*]
[2] Department of Intelligent Mechanical Systems Engineering,
Faculty of Engineering
Kagawa University
Takamatsu,Kagawa,Japan
*corresponding author :
jianguo@tjut.edu.cn

and Jigang Xu[3*]
[3]Unit68709
Qinghai Haidong,810700,China
*corresponding author :
xujigang216@163.com

Abstract **–The spherical robot is mainly used for normal observation of aquaculture biology. The performance of aquatic biological image recognition mainly depends on the feature extraction and the selected classifier. Traditional manual extraction methods often cannot meet actual application requirements, and have problems such as poor accuracy and weak generalization ability. To solve the above problems, a small data set aquatic animal classification model based on convolutional neural network and transfer learning is proposed in the spherical robot. First, the original images of aquatic animals is preprocessed, and the data set is enhanced using the data increment method. Second, The original CNN model is then improved by embedding the SE module and using the triplet loss function to replace the softmax loss function. Finally, Transfer learning a deep pre-trained model of the ImageNet image data set. Training and fitting parameter distributions on aquatic image data sets. Experimental results show that the model optimizes the accuracy of aquatic animal target recognition, and the test accuracy reaches 93.11%. The model has good stability and high precision in aquaculture environment.**

Index Terms –Aquatic animal classification, Deep transfer learning, Few sample learning, Data enhancement.

I. INTRODUCTION

In recent years, as the application of artificial intelligence in the fields of image, machine vision, and speech recognition has matured, deep learning has become one of the research hotspots and mainstream development directions in this field. Spherical robots are mainly used to perform tasks in specific situations. The application of robotic systems to aquaculture in the future will show a diversified and intelligent trend. The application of deep convolutional neural network technology has excellent performance in the field of aquaculture, but the cost of training the model is extremely large, which severely limits the application of this technology in the field of aquaculture[1],[2].

Many research institutions and scholars have conducted a lot of research on target recognition in the field of aquaculture. Maria Teresa Arredondo Waldmeyer have done a lot of research on identifying and locating aquatic animal targets, but they mainly rely on traditional manual extraction of low-level features, such as colour and shape, which often cannot meet the needs of real applications. Jesse Eickholt and Dylan Kelly used the deep learning framework for fish identification, and proposed the deep structure of the deep learning method to mine data in the fish field. Eickholt and Dylan Kelly used LeCun model to complete the study of the recognition of depth convolutional neural networks (DCNN) in fish animals. The above-mentioned deep neural network trained using convolutional neural network achieves good performance. With sufficient image data and annotation quality, the neural network shows excellent performance in image classification. The disadvantage is that the training model is extremely expensive, and there are still challenges in the application of aquaculture [3]-[6].

The purpose of transfer learning is to apply the knowledge learned from the original environment to the method of completing characteristic tasks in the new environment. In addition to the two main tasks of image classification and text classification, this method is also used in the fields of face recognition, food recognition and military target recognition. However, there are fewer studies in the field of aquaculture and lack of specific application examples. [7]-[10].

Due to the nature of the aquaculture target data set, the images collected under public conditions are limited, and the collected image pixels are poor, multiple targets are occluded each other, and single targets are of different types, which will lead to inefficient training and recognition. The data is actually a small data set. In response to this problem, data needs to be enhanced to expand the data set. Goodfellow et al. proposed that the Generative Adversarial Network (GAN) is a generative model that uses generators and discriminators to constantly confront until they are balanced, and finally generate data

978-1-6654-4098-1/21/$31.00 ©2021 IEEE

that conforms to the distribution of real samples. However, the model training process has the problems of instability, difficulty in convergence, and difficulty in training [11]-[14].

We propose a classification model for aquaculture targets based on migration learning and CNN pre-training models, and migrate the big data training model to the new data set rules based on the migration learning method, and use a data enhancement method based on deep convolution to generate a confrontation network. Through the feature extraction of a small number of samples and complete sample reconstruction, to solve the problem of insufficient samples and feature imbalance. It is feasible to use transfer learning and CNN pre-training model for aquaculture target classification training.

II. The Platform of Spherical Robot System

Our spherical robot platform is shown in fig. 1. Spherical robots can complete tasks in specific environments on land and underwater, such as patrols, surveillance, and military operations. In aquaculture, it is impossible to observe aquatic animals in real time through underwater monitoring equipment, and it is difficult to obtain effective breeding data to increase operating income. The spherical robot solves the above-mentioned problems through real-time monitoring of organisms. A monocular camera is installed inside the spherical robot, which can move freely to collect aquatic animal images. Estimate population density by classification of aquatic animals. Obtain real-time breeding data to increase operating income. Therefore, it has important practical significance for the application of spherical robots in the field of aquaculture.

The third part mainly introduces the image classification based on DCNN. The fourth part mainly introduces the experimental data set and processing method. In the fifth part, the classification results using the self-built fish, shrimp, crab and shellfish image verification set show that the optimized classification network effectively improves the accuracy.

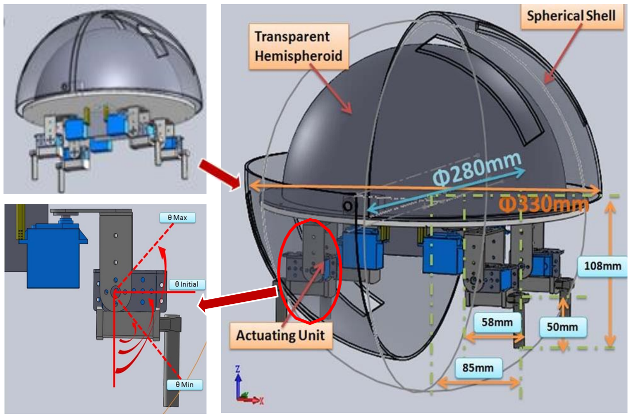


Fig. 1 Spherical amphibious robot platform

III. Image Classification Based on Deep Learning

A. Transfer Learning

With the development of deep learning, many scholars and experts have studied the application of transfer learning based on deep neural network in military target recognition and achieved certain results. Transfer learning is a learning process that utilizes the similarity between data, tasks and models to intelligently apply previously learned knowledge to solve new problems faster or use better solutions, aiming to extract knowledge from the source task and apply knowledge to the target task. We can apply the pre-trained model of big data to the tasks we set, and fine-tune the weight of the pre-trained network through continuous training. In addition, we can adaptively adjust the model according to our task to suit our task[15],[16].

There are three methods of transfer learning :(1) Based on transfer learning, load the trained model and train all network layer parameters on this basis; (2) Feature vectors are extracted. Firstly, the feature vectors of training and test data in the pre-training model are calculated. Then, the pre-training model is abandoned and only the last full-connection layer is trained. (3) Fine tuning. Firstly, all model parameters before the full connection layer of the pre-training model are fixed, and only part of the convolution layer and the full connection layer close to the output end are trained [17]-[19].

B. Resnet Network

When constructing a convolutional network, our general impression is that the deeper the network is, the stronger its expressive ability will be. VGG network is to explore the relationship between the depth of the network and the classification accuracy and found that the network after reaching a certain depth simply increasing layers can't improve the classification accuracy (relative to the shallow network) , and disappeared with the gradient, explosions and degradation risk, such as the main reason is that the redundant network layer learning parameters are not completely the same mapping of input and output are exactly the same (layer). In order to solve the above problems, He Kaiming et al. proposed deep residual network (Resnet), which mainly proposed to replace the full connection layer of the network with residual blocks, and to transform the original identity mapping function into the learning and optimization of the residual function between input and output[20],[21].

For different layers of Resnet network structure, can be divided into two types :(1) based on the building block, the module through including shallow network Resnet34; (2) Based on the bottleneck, the module includes deep network Resnet50/101/152 and even deeper networks. The bottleneck-based design is shown in fig. 2. The bottleneck design directly connects the input data to the third layer through a shortcut connection (the curve part in the figure) without adding additional parameters and computation to complete the equivalent mapping. In addition, the number of parameters is reduced. The first 1×1 convolution layer reduces the 256-dimensional data to 64 dimensions, and then the third 1×1 convolution layer is used to restore the 256-dimensional data.

This process greatly reduces the number of parameters needed to calculate.

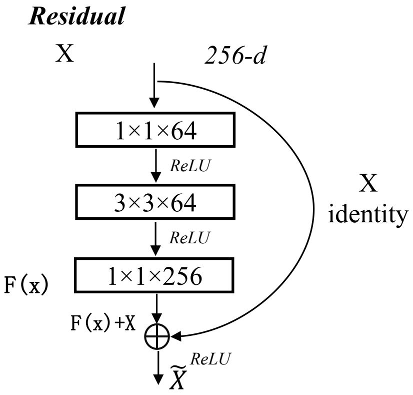


Fig. 2 Bottleneck design

This article chooses the ResNet50 network model, which is mainly composed of three parts: the input convolution layer conv1, a 7×7×64 convolution kernel is responsible for feature extraction, and then 3×3 maximum pooling is used, and then through the residual structure conv2_x, conv3_x, conv4_x, conv5_x, these four residual structures have a total of 3+4+6+3=16 residual layers, and each residual structure is stacked with 3 convolutional layers. Therefore, ResNet50 has a total of 1+16×3=49 convolutional layers.

C. SE Module

Feature compression and excitation network (SENet) is that Hu Jie et al. explicitly model the correlation between CNN feature channels, learn the weights of feature channels, and let the global information of the network enhance useful feature channels and suppress useless feature channels according to the weights. Adaptive calibration of characteristic channels. SENet is a module independent of the neural network structure. Because of its simple structure, there is no need to increase the number of layers or functions, and it can be deployed in the existing convolutional neural network with a small amount of calculation, which can effectively improve the recognition accuracy of the model.

As shown in fig. 3, this is the SE-Resnet network structure that embeds the SE model into the residual network. The SE module consists of two sub-modules: feature compression and excitation. Compression is to use local spatial features to represent global spatial features, and excitation is to learn the weights of feature channels. Compression is to obtain the corresponding value from the feature map of each channel through global flat pooling, and to count the value to represent the global spatial feature. It has the effect of less overfitting, calculation amount and number of parameters.

The flexibility of the SE module is that it can be easily extended to the existing CNN network structure. After the shallower network structure is embedded in the SE module, the recognition accuracy of the model and the generalization ability of different data sets are greatly improved, and the added parameters and the amount of calculation are small. This article chooses to use the embedded SE module in the Resnet-50 network structure, called SE-Resnet-50 [22],[23].

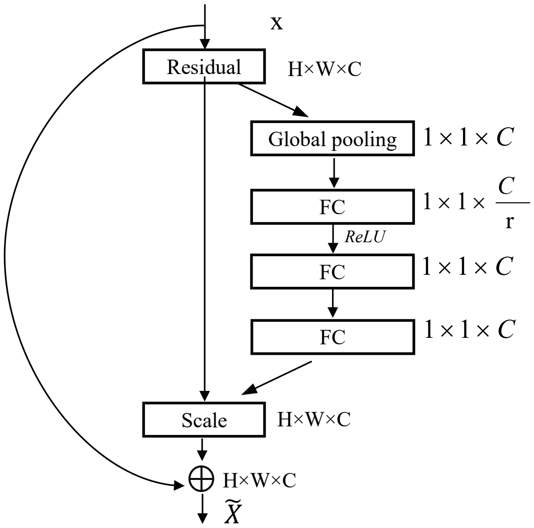


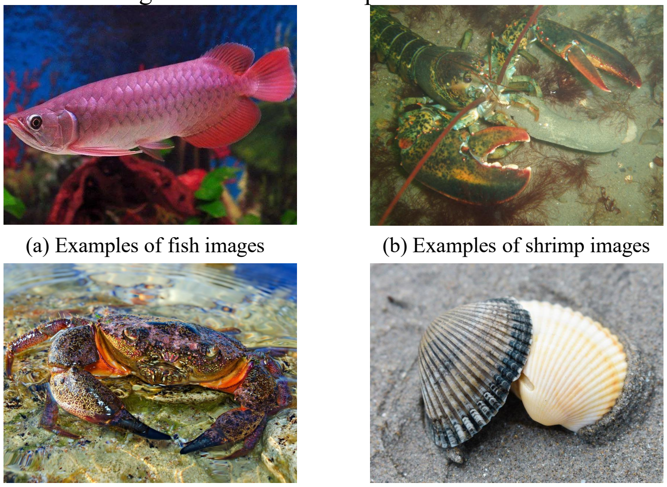
Fig. 3 SE-Resnet module

IV. Experimental Data Set and Processing Method

We use deep learning methods to identify aquatic animals. The data set consists of ImageNet data set and private data set. On this basis, the Resnet pre-training model is constructed, the prepared private data set is input into the network for training labels and other data, and finally the parameter distribution of the fitting aquatic animal data set image is obtained, and then the image is classified and predicted to realize the comparison Detection and classification of aquatic animals.

A. Experimental Data Set Production

The data set is composed of ImageNet data set and private data set. The aquatic animal images in the private data set are obtained by Baidu, Google and other image material websites. If the pixels and resolution of the image are not suitable, the network training error may become larger, so the image needs to be pre-processed. The experiment uses the fish, shrimp, crab and shellfish data set with a total of 1200 pictures, of which the number of pictures in each category is about 300. fig. 4 shows some samples.



(a) Examples of fish images

(b) Examples of shrimp images

(c) Examples of crab images

(d) Examples of shell images

Fig. 4 Examples of aquatic animal images

B. Image Pre-processing

As shown in fig. 5. Before starting the training of the private data set, first use the labelImg software to mark the center area of the aquatic animal image, which is separated from the background and other irrelevant things, and the pending box corresponds to the category name of the target to complete the target labeling. After this treatment, you will get a better training effect. After marking, an xml file will be generated. The image information contained in the file has four coordinate points, the image category name, and the file storage path. When training the network, input the xml file into the network, so there is no need to read the entire image for detection, which can effectively reduce the overall parameters of the network, thereby avoiding the influence of irrelevant background and objects on the final recognition result.

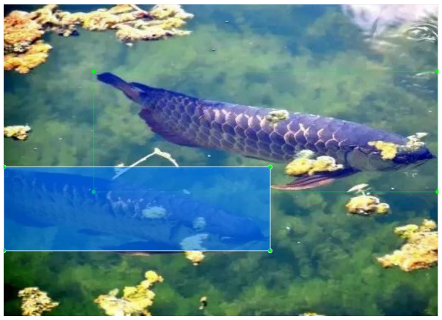


Fig. 5 Annotate dataset images

C. Image Enhancement Method

The establishment of aquatic animal image data set needs to consider the diversity of the target and the complexity of the environment in which it is located. At the same time, it must also consider the impact of the target in different collection, light, shelter, and water quality conditions. In order to ensure the richness of training samples, improve the generalization ability of the model, and avoid over-fitting, it is necessary to use data enhancement methods to expand the number of samples.

The number of samples before and after data expansion is shown in Table I. According to different theories that different samples may appear in the real environment, a data enhancement method based on Deep Convolution Generative Adversarial Network (DCGAN) is adopted. The aquatic animals collected in this article are divided into four categories.

The image data enhancement method we use is a deep convolutional generation confrontation network(DCGAN).

TABLE I

STATISTICS OF AQUATIC ANIMAL DATA SETS

Category	Original	DCGAN
fish	330	1650
shrimp	240	960
crab	270	1350
shell	360	1800
total	1200	5760

we normalized the images so that the size of all sample images was unified to 224×224, and saved the processed data as an Img format file, which was then used for sample labels for subsequent deep network training.

D. Network Model Design

Convolutional neural networks use loss functions to evaluate the training model. The role of the loss function is to estimate the degree of difference between the actual value of the model and the predicted value. Cross-entropy (Logloss) Softmax is the most commonly used loss function for multi-class neural network output, and its output represents the relative probability between different categories. The characteristic is that it is easy to separate the characteristics of different categories, and it is not easy to separate the characteristics of similar categories. Due to the complex environment of aquatic animals, it is easy to cause large intra-class distances and small inter-class distances, resulting in poor discrimination between classes. Moreover, the types and quantities of samples required for aquatic animals are particularly large, which makes the use of feature vectors to distinguish aquatic products Animals become more difficult[24].

In order to solve the above problems, Florian Schroff and others proposed the Triplet loss function instead of the Softmax loss function method, which mainly trains samples with small differences, and defines the ternary including Anchor example, Positive example, and Negative example. Group, which represents the distance between samples of the same type and samples of different types. By comparing three images to define the triple loss, it reduces the intra-class distance and increases the inter-class distance. The calculation formula of the Triplet loss function is as follows[25],[26]:

$$L_{\min} = \sum_{i}^{N} [\| f(x_i^a) - f(x_i^p) \|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + a]_+ \quad (1)$$

In the formula (1), x_i^a is i anchor, x_i^p is i positive image, x_i^n is negative images, a is the offset. $f(x_i^a)$ Means to minimize the anchor point, $f(x_i^p)$ Represents the distance between the anchor point and the positive sample of the same kind, $f(x_i^n)$ Represents the distance between the anchor point and the heterogeneous negative sample. We expect that the distance between $f(x_i^a)$ and $f(x_i^p)$ is small, the distance between $f(x_i^a)$ and $f(x_i^n)$ is large. we make the features of similar samples as close as possible in the European space, while the features of heterogeneous samples are as far away as possible in the European space. To prevent sample features from gathering in a small space, there is at least one offset a between the two distances.

V. EXPERIMENTAL TEST AND RESULT ANALYSIS

A. Experimental Configuration

In terms of data sets. We divide the aquatic animal data set into three parts according to the ratio of 7:2:1: training set, test set, and validation set. The image size is pre-processed to (224, 224). In terms of hyperparameters. The initial learning rate is set to 0.001, the learning rate is reduced by 1/10 every

3 periods, and the momentum is 0.5. In terms of pre-training models. Using ResNet50 and SE-ResNet50 as the basic network, using the ImageNet data set for pre-training, initializing the weights of the original neural network, and obtaining a pre-training model for migration learning.

In this experiment, we used ResNet50 and SE-ResNet50 as the basic network to conduct two sets of experiments: (1) The migration learning target classification of the residual network, combined with the Softmax loss function, is expressed as ResNet50-Softmax and SE-ResNet50- Softmax. (2) The residual network combined with the Triplet loss function is expressed as ResNet50-Triplet and SE-ResNet50-Tripletrespectively.

Residual network using Softmax loss function. Perform migration fine-tuning training on the aquatic animal data set. Since the residual network does not have an FC layer, the original layer is replaced with a feature layer (2048 dimensions), the convolution kernel uses 1×1, and the output is a 5-dimensional convolution layer. Residual network using Triplet loss function. Replace the Softmax loss function in the pre-training model with the Triplet loss function, and then fine-tune the aquatic animal data set.

Our experiments use accuracy and precision as the evaluation criteria of the model. The accuracy uses Top-1 Accuracy and Top-5 Accuracy standard. To classify the entire sample, the former selects the accuracy rate of the predicted category with the highest predicted probability in the classification result and the actual category. The latter refers to the accuracy of the top 5 predicted categories and the true categories.

B. Experimental Results and Analysis

We tested the trained model with a test set of aquatic animals. The test results are shown in Table II. Compared with the classic ResNet50-Softmaxnetwork, SE-ResNet50-Softmax has increased the accuracy and average accuracy of Top-1 by 1.36% and 1.43%, respectively. After the training network is added to the SE module, the efficiency of aquatic animal classification is effectively improved. This is because neural networks use global information to enhance useful information while suppressing useless information. First, the response on the two-dimensional feature channel is compressed into a real number, which can have the same global receptive field as before compression, and shallow information can also represent deep information, converging and expanding the receptive field. In addition, the correlation of each characteristic channel is expressed through modeling design parameters. Finally, adaptively recalibrate the importance of each feature channel after feature selection. Thereby increasing the generalization ability of the network and the ability to recognize the target.

It can be seen from Table II, the basic ResNet50-Softmax network structure, Top-1 Achieved 82.56% accuracy rate. Top-5 Achieved 100% accuracy rate and Mean accuracy Achieved 88.50% accuracy rate.

The loss function affects the recognition accuracy of the network model. Using the classic ResNet50-Softmax network has the lowest performance in Top-1 and average accuracy. After replacing the loss function of the network with the Triplet loss function, the Top-1 and average accuracy rates were increased by 2.08% and 3.03%, respectively.

TABLE II

TOP1AND TOP5 CLASSIFICATION ACCURACY

Model	Top-1(%)	Top-5(%)	Mean accuracy(%)
ResNet50-Softmax	82.56	100.00	88.50
SE-ResNet50-Softmax	83.92	100.00	89.93
ResNet50-Triplet	84.64	100.00	91.53
SE-ResNet50-Triplet	87.27	100.00	93.84

After the two loss function networks were added to the SE module at the same time, the recognition accuracy rate was significantly improved. Increased by 3.35% and 3.91% respectively. The experimental results show that adding the Triplet loss function to the network can better complete the aquatic animal detection and classification tasks than the Softmax loss function, and is more suitable for the high-precision and high-accuracy requirements of aquatic animal classification tasks.

VI. CONCLUSIONS AND FUTURE WORK

According to the classification requirements of aquatic animal detection, this paper proposes a small data set aquatic animal classification and detection model based on transfer learning and convolutional neural network. A private data set was made and the original image was pre-trained to create a small data set of aquatic animals including four categories of fish, shrimp, crabs and shellfish. First, the SE module is embedded in the residual network, and the maximum loss function is replaced with a triple loss function, and design the SE-Resnet-T model. The experimental results show that the model optimizes the classification problem of aquatic animal detection, and the test accuracy rate reaches 93.84%, which can better complete the aquatic animal detection and classification tasks. The model has good stability and high recognition accuracy in complex environments.

ACKNOWLEDGMENTS

This research is supported by National Natural Science Foundation of China (61703305), Key Research Program of the Natural Science Foundation of Tianjin (18JCZDJC38500) and Innovative Cooperation Project of Tianjin Scientific and Technological (18PTZWHZ00090).

REFERENCE

[1] Yong Luo, Yonggang Wen, Lingyu Duan and Dacheng Tao, "Transfer Metric Learning: Algorithms, Applications and Outlooks", *Journal of Latex Class Files*, 2018.

[2] DE. Rumelhart, G E. Hinton, R J. Williams, "Learning Representations by Back Propagating Errors", *Nature*, 1986.

[3] Maria Teresa Arredondo Waldmeyer, "A Novel Framework for Automated image set preparation for moving objects in under water videos", *International journal of computational intelligence research*, Vol.14 No.12 pp:1041-1060,2018.

[4] Pawan K. Ajmera, Harsh Sinha and Vinayak Awasthi, "Audio Classification using Braided Convolutional Neural Networks", *IET Signal Processing*, Vol.14, No.7, pp:448-454,2020.

[5] A. Marburg and K. Bigham, "Deep learning for benthic fauna identification", *IEEE*, Vol.9, No.12, pp:1-5, 2016.

[6] LeCun, Yann, Bengio, Yoshua, Hinton and Geoffrey, “Deep learning”, *Nature*, 2015.

[7] Xin Sun, Hongwei Xi, Junyu Dong and Huiyu Zhou, “Few-Shot Learning for Domain-Specific Fine-Grained Image Classification”, *IEEE Transactions on Industrial Electronics. IEEE*, Vo19, No.12, pp:1-1, 2020.

[8] Xueting Zhang, Flood Sung, YutingQiang and Yongxin Yang, “Deep Comparison: Relation Columns for Few-Shot Learning”,*2020 International Joint Conference on Neural Networks*, Vol.10, No.5, pp:165-172,2020.

[9] YuxiangXie, Hua Xu, Congcong Yang and Kai Gao, “Multi-Channel Convolutional Neural Networks with Adversarial Training for Few-Shot Relation Classification (Student Abstract)”, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol.34, No.10,pp:13967-13968, 2020.

[10] Chen Long, Zhang Feng and Jiang Sheng. “A typical military target recognition method based on deep forest learning model under small sample conditions”,*Journal of China Academy of Electronics*, Vol.14, No.3, pp:232−237,2019.

[11] Y J. Park, et al, “Insect Classification Using Squeeze-and-Excitation and Attention Modulesa Benchmark Study”, *2019 IEEE International Conference on Image Processing. IEEE*, 2019.

[12] L. Taylor, G. Nitschke. “Improving Deep Learning with Generic Data Augmentation”, *2020 IEEE Symposium Series on Computational Intelligence* , Vol.7, No.3, pp:1542-1547,2020.

[13] Tero Karras, et al, “Analyzing and Improving the Image Quality of StyleGAN”, *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE* ,2020.

[14] L. J. Ratliff, S. A. Burden and S. S. Sastry, “Characterization and computation of local Nash equilibria in continuous games”,*2019 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello. IL*, pp:917-924, 2019.

[15] Junlin Hu, Jiwen Lu and Yap-Peng Tan, “Deep transfer metric learning”,*2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*, 2020.

[16] F.Z. Zhuang, Ping Luo, Qing He and Zhongzhi Shi, “Survey on transfer learning research”, *Journal of Software*, Vol.26, pp:26-39, 2019.

[17] S. Rodionov, et al, “Improving Deep Models of Person Re-identification for Cross-Dataset Usage”, *Artificial Intelligence Applications and Innovations*, 2018.

[18] Dong Yi, Zhen Lei, Shengcai Liao and Stan Z. Li, “Deep Metric Learning for Person Re-identification”, *International Conference on Pattern Recognition. IEEE Computer Society*, 2020.

[19] Xuhong Wei,Yefei Chen and JianboSu, “Domain Adaptation via Identical Distribution Across Models and Tasks”, *Lecture Notes in Computer Science, Springer, Cham*,Vol.11301,2019.

[20] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, *Computer Science*, 2019.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. “Deep residual learning for image recognition”,*2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*, 2019.

[22] Yushi Chen, Zhouhan Lin, Xing Zhao and Gang Wang, “Deep LearningBased Classification of Hyperspectral Data”, *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, Vol.7, No.6, pp:2094-2107,2019.

[23] Jie Hu, Li Shen, Gang Sun and Samuel Albanie, “Squeeze-and-Excitation Networks”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp:99,2020.

[24] Weiyang Liu,Yandong Wen,Zhiding Yu and Meng Yang, “Large-Margin Softmax Loss for Convolutional Neural Networks”, *Journal of Machine Learning Research*, 2019.

[25] Florian Schroff, Dmitry Kalenichenko and James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, *IEEE*, 2019.

[26] Fuqing Zhu, Xiangwei Kong, Liang Zheng and Haiyan Fu, “Part-Based Deep Hashing for Large-Scale Person Re-Identification”, *IEEE Transactions on Image Processing*, Vol.26, No.10, pp:4806-4817, 2020.