Data Augmentation Using Image-to-image Translation for Tongue Coating Thickness Classification with Imbalanced Data

1st Mingxuan Liu
Department of biomedical engineering
Tsinghua University
Beijing China
liumx19@mails.tsinghua.edu.cn

2nd Yunrui Jiao
School of Integrated Circuits
Tsinghua University
Beijing China
jyr19@mails.tsinghua.edu.cn

3rd Hongyu Gu
School of Integrated Circuits
Tsinghua University
Beijing China
ghy19@mails.tsinghua.edu.cn

4th Jingqiao Lu
Health Testing Technique And Equipment Research Center
Huangpu Joint Innovation Institute Of Chinese Medicine
Guangzhou China
lujingqiao@jiicm.org.cn

5th Hong Chen
School of Integrated Circuits
Tsinghua University
Beijing China
hongchen@tsinghua.edu.cn

2022 IEEE Biomedical Circuits and Systems Conference (BioCAS) | 978-1-6654-6917-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/BioCAS54905.2022.9948645

Abstract—**Tongue diagnosis is widely used in traditional Chinese medicine diagnosis. The classification of tongue coating thickness is one of the most important tasks in tongue diagnosis. However, data imbalance imposes challenges when using deep learning methods for tongue coating thickness classification. In this paper, we propose a data augmentation method using image-to-image translation to solve the data imbalance problem. First, we use an image-to-image translation model based on generative adversarial networks (GANs) to translate thick and thin tongue coating images into each other, then we train the classification model using synthetic images together with real images. Finally, the trained classification model is used to classify the thickness of tongue coating. With our data augmentation method, the classification performance yields 0.92 accuracy and 0.922 F1-score, which is 3.37% and 3.95% higher than that with re-sampling method respectively.**

Keywords—TCM, tongue diagnosis, data augmentation, image classification, image-to-image translation

I. INTRODUCTION

Tongue diagnosis is an important step in traditional Chinese medicine (TCM). The color and coating of the tongue can help to understand the physiological mechanisms of the body and the pathology of diseases[1]. For example, Tsung-Chieh Lee et al. found that metabolic syndrome (MetS) can be diagnosed by tongue coating analysis and heart rate variability devices[2]. In a recent study, Zhichun Wang et al. found that coronavirus disease 2019 (COVID-19) subjects had pathological tongue coating patterns that are associated with inflammatory responses[3].

However, the judgment of tongue features by TCM practitioners usually depends on personal knowledge and experience, which has been criticized as lacking objectivity[4]. Many tongue feature classification algorithms based on deep learning and machine learning have been proposed to assist doctors in diagnosis. One of the main challenges in medical image classification is how to deal with data imbalance[5]. Classification models will typically over-classify the majority class when the training data is imbalanced, which is because its prior probability is higher. As a result, data belonging to the minority class is more likely to be misclassified[6]. Generally, the classification of tongue features also has the problem of data imbalance.

Many efforts have been made to solve data imbalance, including three main solutions: class re-balancing, module improvement, and information augmentation[7]. Class re-balancing is the mainstream paradigm for solving such problems, such as re-sampling [8], cost-sensitive learning [9], and logit adjustment [10]. Module improvement includes four main approaches: representation learning[11], classifier design, decoupled training[12], and ensemble learning[13]. Information augmentation can be divided into two types: transfer learning[14] and data augmentation.

There have been several studies adopting the above approaches to solve the problem in tongue image classification. In [15], the authors performed deep transfer learning training on GoogleNet and ResNet and obtained better classification performance than that of the model without transfer learning. Authors in [16] improved the tongue coating recognition model using transfer learning techniques and applied it to COVID-19 diagnosis. Other researchers adopt model improvement to improve classification performance. In [17], tongue feature classification was used for somatic recognition, and a complexity perception (CP) classification method is proposed to alleviate the negative impact of uneven distribution of tongue images on classifier performance. Multiple-instance learning (MIL) was used in [18] to resolve inconsistent performance due to the varying size or location of the tongue coating region. Multi-task joint learning model (MTL) was adopted for tongue segmentation and feature classification, and it was experimentally demonstrated to achieve balanced improvements in accuracy and recall for each class[19].

Although above methods bring some improvements to the classification of tongue features from imbalanced datasets in terms of accuracy, they are usually highly sensitive to hyperparameters or have complex training procedures[20]. Different from the other methods, the data augmentation method only needs simple refinements to the training

This work is supported by the National Science and Technology Major Project from Minister of Science and Technology, China (Grant No. 2018AAA0103100), Guangzhou Foshan Science and Technology Innovation Project (No. 2020001005585), and National Natural Science Foundation of China (No. 92164110 and U19B2041), partly supported by Beijing Engineering Research Center (No. BG0149) and Tsinghua National Laboratory for Information Science and Technology (No. 042003266).

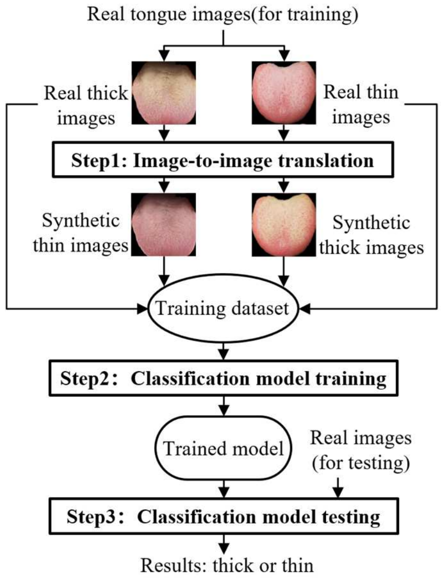

Fig. 1.Overview of the proposed method.

procedure. In this paper, a data augmentation method based on image-to-image translation is proposed as shown in Fig. 1, from which we can see that the proposed method consists of three steps: the first step is image-to-image translation, with which we translate thick and thin tongue coating images into each other, thereby obtaining synthetic thick and thin tongue coating images. In the second step, the synthetic images are used as the training dataset together with the real images to train a classification model. In the final step, the trained model is used to classify the tongue coating thickness.

II. Proposed Methods

A. Generating synthetic tongue images

Generally, there are two types of GAN-based data augmentation methods. The first method is to train a GAN with the data only from the target class and mapping is obtained between random noise and images of the target class[21,22,33]. But when the data is imbalanced, the data in the minority class is too little to be used for training a GAN [24]. Another method is to train a conditional GAN or its variants using data from all classes[25,26]. Taking ACGAN[27] as an example, the generator of ACGAN generates images for each class during training, and the discriminator outputs the results of the image: real or fake and its class. However, when the data is imbalanced, the minority class images are easily recognized as fake by the discriminator. As a result, it is difficult to find a generated image that can be judged as a both real and minority class image with the method. In order to generate minority class images for data augmentation successfully, we first adopt a GAN-based image-to-image translation model. The model is used for the image to image mappings rather than the random noise to images mapping.

Now we describe the GAN-based image-to-image translation in detail. X and Y are two image domains. The translation aims to learn the mappings $f_{x\to y}$ and $f_{y\to x}$ with marginals $p(X)$ and $p(Y)$. Most frameworks (such as CycleGAN[28]) use a combination of the encoder E_x (resp. E_y) and the generator $G_{x\to y}$ (resp. $G_{y\to x}$) to implement image translation. The output is $G_{x\to y}(E_x(x))$(resp. $G_{y\to x}(E_y(y))$). In order to train the generator, we need a discriminator D_y(resp. D_x) to classify the output into real and synthetic images and obtain the loss of the discriminator to optimize the generator.

The image-to-image translation model generally has two encoders, two generators and two discriminators, so it has a large number of parameters which makes the training process takes a long time. We used the no independent component encoding GAN (NICE-GAN)[29] to obtain a more compact model, whose encoder is part of the discriminator. The flowchart of NICE-GAN is shown in Fig. 2, from which we can see that the discriminator D_y(resp. D_x) is composed of two parts: the encoder E_x^D(resp. E_y^D) and the classifier C_x^D(resp. C_y^D). The encoder E_x^D(resp. E_y^D) replaces the original encoder E_x(resp. E_y) to obtain the output $f_{x\to y}(x)=G_{x\to y}(E_x^D(x))$(resp. $f_{y\to x}(y)=G_{y\to x}(E_y^D(y))$). As a result, the model is simplified and the training time is shorter.

In NICE-GAN a multi-scale discriminator is used to provide more information for the training process, but it makes the network more complex. We replaced the multi-scale discriminator with a single discriminator to make the architecture more concise. The customized NICE-GAN architecture for this study is shown in Table I. The four numbers in Conv(64,4,2,1) stand for: number of channels, kernel size, side size and padding size respectively; SN means the spectral normalization; LR represents the LeakyReLU activation function and the slope is set to 0.2; LIN is layer-instance normalization; and FC is the fully connected layer, followed by the number of output channels.

We used all images from the training dataset and resized them to 512 × 512 when training the NICE-GAN network. The loss function in [29] was used and optimized by the Adam optimizer with the learning rate set to 0.0001 and (β1,b2) set to (0.5,0.999). The model was trained on NVIDIA RTX 3090 GPUs using random level flipping as well as random cropping for data augmentation training. Set N_a and N_b as the number of thick and thin tongue coating images respectively, after the training we can obtain N_b synthetic thick images and N_a synthetic thin images.

We evaluate the quality of synthetic images by measuring Inception Score (IS)[30], Fréchet Inception Distance (FID)[31] and Kernel Inception Distance (KID)[32] at different epoch to find the optimal training epoch. We will finish the training to save time when they are stable.

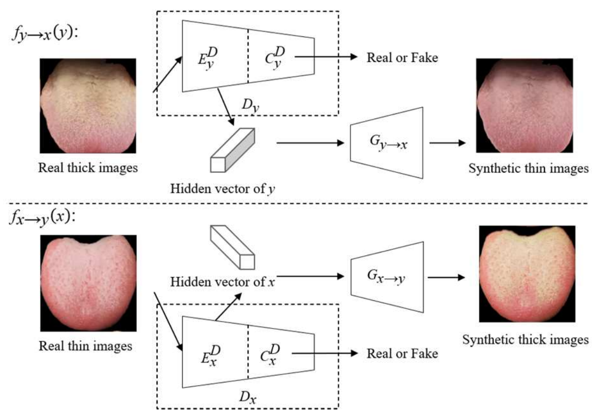

Fig. 2.Flowchart illustration of NICE-GAN

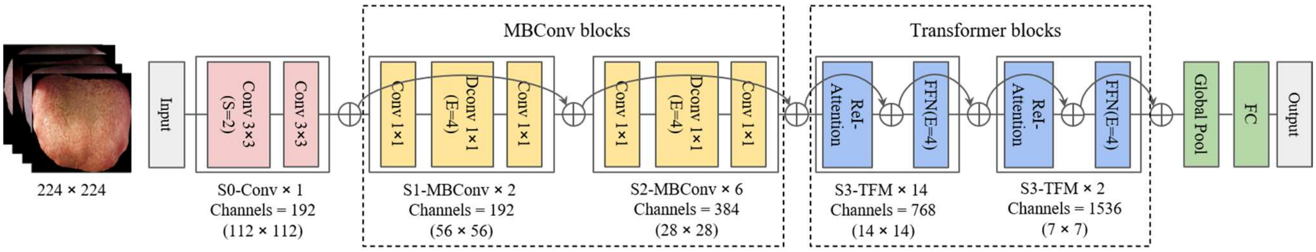

Fig. 3.Schematic of CoAtNet architecture for tongue coating thickness classification

TABLE I. ARCHITECTURE OF NICE-GAN

Discriminator	***Input to output***	***Layer***
Encoder	(512,512,3) to (256,256,64)	Conv(64,4,2,1),SN,LR
	(256,256,64) to (128,128,128)	Conv(128,4,2,1),SN,LR
Down-sampling0	(128,128,128) to (64,64,256)	Conv(256,4,2,1),SN,LR
Down-sampling1	(64,64,256) to (32,32,512)	Conv(512,4,2,1),SN,LR
	(32,32,512) to (16,16,1024)	Conv(1024,4,2,1),SN,LR
Classifier	(16,16,1024) to (15,15,2048)	Conv(2048,4,1,1),SN,LR
	(15,15,2048) to (14,14,1)	Conv(1,4,1,1),SN
Generator	***Input to output***	***Layer***
Sampling	(128,128,128) to (128,128,256)	Conv(256,3,1,1),LIN,ReLU
γAdaLIN, βAdaLIN	(128,128,256) to (1,1,256)	Global Average Pooling
	(1,1,256) to (1,1,256)	[FC(256),ReLU]×3
Bottleneck	(128,128,256) to (128,128,256)	[AdaResBloack(256,3,1,1), AdaLIN,ReLU]×6
Up-sampling	(128,128,256) to (256,256,128)	Sub-pixel-Conv(128,3,1,1), LIN,ReLU
	(256,256,128) to (512,512,64)	Sub-pixel-Conv(64,3,1,1), LIN,ReLU
	(512,512,64) to (512,512,3)	Conv(3,7,1,3),Tanh

B. *Classification model training*

Ninety percent of our synthetic images are randomly selected and added to the training dataset, so the number of thick tongue coating images is $N_a + 0.9N_b$ and the number of thin is $N_b + 0.9N_a$ in the training dataset. Convolutional neural networks (ConvNets) are often used for image classification. Moreover, with self-attention models, better performance in natural language processing has been achieved[33]. The self-attention model is also used for image classification tasks because it has a higher capacity at scale and faster convergence compared to ConvNets. But its performance still falls behind the state-of-the-art ConvNet in the low data regime[34]. We designed a combined model of convolution and self-attention to take advantage of them. And we use it for the classification of tongue coating thickness.

As discussed in [34], CoAtNet is a newly proposed model for classification, which naturally unifies depth-wise convolution and self-attention through simple relative attention, and generalizes the model by stacking convolution and attention layers vertically in a principled way. In order to make the convergence faster and ensure high accuracy of the classification at the same time, we designed a CoAtNet model applicable to tongue coating thickness classification by experiments. The designed CoAtNet architecture can be seen in Fig. 3, which consists of convolutional blocks and transformer blocks. S0 includes two convolutional layers, S1 and S2 are MBConv blocks[35] with the channels set to 192 and 384 respectively. S3 and S4 are transformers blocks with the channels set to 768 and 1536 respectively, and they have the same configurations as [36].

We adopted Label-Distribution-Aware Margin Loss (LDAMLoss)[37] function to train the CoAtNet learning parameters for getting the best recognition performance, which is illustrated as follows:

$$L_{LDAM}\left((x,y);f\right) = -log\,\frac{e^{z_y - D_y}}{e^{z_y - D_y} + \sum_{j \neq y} e^{z_j}} \tag{1}$$

$$D_j = \frac{C}{n_j^{1/4}} \text{ for } j \in \{1,\ldots,k\} \tag{2}$$

where D_j indicates a class-dependent margin, y is the true label of input x, j indicates a class, n_j stands for the number of samples of class j, and C is a hyperparameter. For z_j , let f be a model, (x,y) be an input example, and $z_j = f(x)_j$ indicates the j-th output of the model for the j-th class.

We used the Adam optimizer to train the customized CoAtNet on NVIDIA RTX 3090 GPUs for all training dataset. We first randomly cropped the 512×512-sized image by 0.8 times its area and then resized it to 256×256. Then, we adopted the random rotation function, set the degrees to 15, finally flipped the image randomly horizontally and resized it to 224×224. The training epoch was set to 500.

III. RESULTS

A. *Tongue thickness data*

The tongue images were collected from several TCM hospitals and labeled by senior doctors. We have got 1,010 tongue images, of which 836 and 174 were labeled as thin and thick tongue coating respectively. The dataset is very


Fig. 4.Examples of Synthetic images, from top to bottom: real thick, synthetic thin from real thick, real thin, synthetic thick from real thin.

imbalanced. We used an automatic tongue image segmentation algorithm[38] to pre-process the collected images. To obtain a credible accuracy, we randomly selected 50 thick and 50 thin tongue coating images as the testing dataset, and the remaining images as the training dataset for NICE-GAN, then we trained the CoAtNet with the synthetic images and the real images, and finally we tested the classification performance on the testing dataset.

B. Evaluation of the Synthetic images from NICE-GAN

The images synthesized by NICE-GAN are shown in Fig. 4, from which we can see that the tongue shape in synthetic images (row 2 and 4 in Fig. 4) is the same as that in original images (row 1 and 3 in Fig. 4), only the thickness of the tongue coating is transformed. We test the IS, FID and KID of all synthetic images under different epoch to quantitatively evaluate the quality of synthetic images and find the best training epoch. The experimental results are shown in Fig. 5. As the smaller FID and KID metrics means the higher quality of synthetic images, and the higher IS metrics indicates the

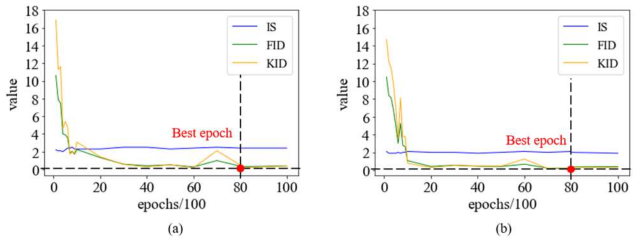

Fig. 6. The IS, FID, and KID of synthetic thick tongue coating images (a) and thin (b) with the increase of training epochs. increase of training epochs.

higher quality, it can be noticed that the best quality of synthetic images is obtained when the training epoch is 8000. Compared with the default epoch of 100,000 suggested by [29], our training time is saved significantly. Besides, we find that the IS metric remains a constant during the training process, which is because the IS metric only reflects the diversity of synthetic images themselves not the similarity between synthetic samples and training samples.

C. Evaluation of the Proposed Data Augmentation

To verify our method, we compare the classification performance of the model with other methods in terms of four metrics: accuracy (ACC), sensitivity in (3), specificity in (4), and F1-score in (6). For the binary classification problem, we define thick tongue coating as positive, and TP, FP, TN, FN stand for the numbers of true positives, false positives, true negatives and false negatives described in (3)-(6) respectively. We note that the sensitivity and specificity are also the recall of positives and negatives respectively.

$$Sensitivity = TP \,/\, (TP + FN) \tag{3}$$

$$Specificity = TN \,/\, (TN + FP) \tag{4}$$

$$precision = TP \,/\, (TP + FP) \tag{5}$$

$$F1\text{-}score = \frac{2 \times precision \times Sensitivity}{precision + Sensitivity} \tag{6}$$

We compare our data augmentation method with the method of re-sampling, which is widely used to resolve the problem of data imbalance. With the re-sampling way, each class has an equal probability of being selected [8]. We test two tongue coating thickness classification models (ConvNets models) in [39] and [40] for comparison.

The performance comparison results are shown in Table II, from which we find that with the re-sampling method, the sensitivity of the designed CoAtNet model is much lower than the specificity because there are more thin tongue coating images in the training dataset. That means the re-sampling method fails to deal with the data imbalance. In contrast, with our method, the sensitivity increases from 0.86 to 0.94, the specificity decreases slightly (from 0.92 to 0.90), and the accuracy increases from 0.89 to 0.92. In addition, when using

TABLE II. PERFORMANCE COMPARISON

Method / Model	Re-sampling			Our work		
	Sens.	***Spec.***	***ACC***	***Sens.***	***Spec.***	***ACC***
Designed CoAtNet	0.86	0.92	0.89	0.94	0.90	0.92
Ref.[39]	0.74	0.70	0.72	0.62	0.86	0.74
Ref.[40]	0.46	0.98	0.72	0.82	0.78	0.80

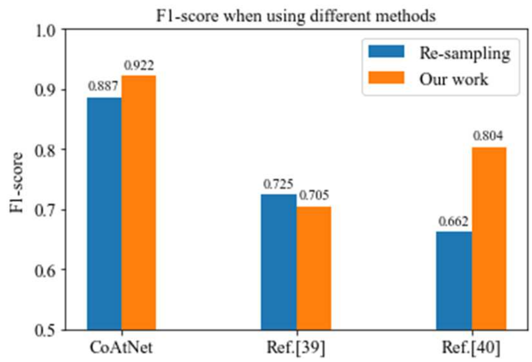

Fig. 5.F1-score of two methods

the classifiers in [39] and [40], our data augmentation method improves the accuracy by 2.78% and 11.11% respectively compared with the re-sampling method.

Besides, as the F1-score takes into account both the accuracy and recall of the classification model, it can be used to evaluate methods. The comparison results of the F1-score are shown in Fig. 6, from which we can see that with our method the F1-score is improved from 0.887 to 0.992 with CoAtNet and from 0.662 to 0.804 with the model in [40] respectively. But F1-score is reduced by 2.76% when using the model in [39].

IV. CONCLUSIONS

In this work, we proposed a method that uses a GAN-based image-to-image transformation for data augmentation to improve classification performance with imbalanced data. In the tongue coating thickness classification task, with our method the accuracy and F1-score are successfully improved by 3.37% and 3.95% respectively when using CoAtNet. In addition, with the CoAtNet model, our method has achieved the highest accuracy of 0.92 and F1-score of 0.922. In future work, more features of the tongue will be classified with our method.

REFERENCES

[1] Dong Zhang, Junhua Zhang, Zheng Wang and Meijun Sun. “Tongue colour and coating prediction in traditional Chinese medicine based on visible hyperspectral imaging.” IET Image Processing 13.12 (2019): 2265-2270.

[2] Tsung-Chieh Lee, Lun-Chien Lo, and Fang-Chen Wu. “Traditional Chinese medicine for metabolic syndrome via TCM pattern differentiation: Tongue diagnosis for predictor.” Evidence-Based Complementary and Alternative Medicine 2016 (2016).

[3] Wangzhi Chun, et al. “Tongue Coating in COVID-19 Patients: A Case-Control Study.” medRxiv (2022).

[4] Lun-chien Lo, Yung-Fu Chen, Wen-Jiuan Chen, Tsung-Lin Cheng and John Y.Chiang. “The study on the agreement between automatic tongue diagnosis system and traditional chinese medicine practitioners.” Evidence-Based Complementary and Alternative Medicine 2012 (2012).

[5] H. Greenspan, B. van Ginneken, and R. M. Sum-mers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting newtechnique.” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1153–1159, May 2016

[6] Johnson, Justin M., and Taghi M. Khoshgoftaar. “Survey on deep learning with class imbalance.” Journal of Big Data 6.1 (2019): 1-54.

[7] Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, Jiashi Feng. “Deep long-tailed learning: A survey.” arXiv preprint arXiv:2110.04596 (2021).

[8] Estabrooks, Andrew, Taeho Jo, and Nathalie Japkowicz. “A multiple resampling method for learning from imbalanced data sets.” Computational intelligence 20.1 (2004): 18-36.

[9] Elkan, Charles. “The foundations of cost-sensitive learning.” International joint conference on artificial intelligence. Vol. 17. No. 1. Lawrence Erlbaum Associates Ltd, 2001.

[10] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit and Sanjiv Kumar. “Long-tail learning via logit adjustment.” arXiv preprint arXiv:2007.07314 (2020).

[11] Bengio, Yoshua, Aaron Courville, and Pascal Vincent. “Representation learning: A review and new perspectives.” IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1798-1828.

[12] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng and Y. Kalantidis, “Decoupling representation and classifier for long-tailed recognition.” In International Conference on Learning Representations, 2020.

[13] Sagi, Omer, and Lior Rokach. “Ensemble learning: A survey.” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.4 (2018): e1249.

[14] Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang. “A survey of transfer learning.” Journal of Big data 3.1 (2016): 1-40.

[15] Chao Song, Bin Wang and Jiatuo Xu. “Research on tongue feature classification method based on deep migration learning.” Computer Engineering and Science 43.08(2021):1488-1496. (in Chinese)

[16] Wang, Xu, et al. “Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in noninvasive ethnopharmacological evaluation.” Journal of Ethnopharmacology 285 (2022): 114905.

[17] Jiajiong Ma, Guihua Wen, Changjun Wang and Lijun Jiang. “Complexity perception classification method for tongue constitution recognition.” Artificial intelligence in medicine 96 (2019): 123-133.

[18] Y. Tang, Y. Sun, J. Y. Chiang and X. Li, “Research on Multiple-Instance Learning for Tongue Coating Classification.” In IEEE Access, vol. 9, pp. 66361-66370, 2021, doi: 10.1109/ACCESS.2021.3076604.

[19] Qiang Xu, et al. “Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network.” IEEE journal of biomedical and health informatics 24.9 (2020): 2481-2489.

[20] Yongshun Zhang, XiuShen Wei, Boyan Zhou, Jianxin Wu. “Bag of tricks for long-tailed visual recognition with deep convolutional neural networks.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 4. 2021.

[21] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger and Hayit Greenspan. “Synthetic data augmentation using GAN for improved liver lesion classification.” 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE, 2018.

[22] Levine, Adrian B., et al. “Synthesis of diagnostic quality cancer pathology images by generative adversarial networks.” The Journal of pathology 252.2 (2020): 178-188.

[23] Devansh Bisla, Anna Choromanska, Russell S. Berman, Jennifer A. Stein and David Polsky. “Towards automated melanoma detection with deep learning: Data purification and augmentation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[24] Gurumurthy, Swaminathan, Ravi Kiran Sarvadevabhatla, and R. Venkatesh Babu. “Deligan: Generative adversarial networks for diverse and limited data.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[25] Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman and Plácido Rogerio Pinheiro “Covidgan: data augmentation using auxiliary classifier gan for improved covid-19 detection.” Ieee Access 8 (2020): 91916-91923.

[26] Jordan J.Bird, Chloe M.Barnes, Luis J.Mansoc, Anikó Ekárt and Diego R.Fariac. “Fruit quality and defect image classification with conditional GAN data augmentation.” Scientia Horticulturae 293 (2022): 110684.

[27] Odena, Augustus, Christopher Olah, and Jonathon Shlens. “Conditional image synthesis with auxiliary classifier gans.” International conference on machine learning. PMLR, 2017.

[28] Junyan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. “Unpaired image-to-image translation using cycle-consistent adversarial networks.” Proceedings of the IEEE international conference on computer vision. 2017.

[29] Runfa Chen, Wenbing Huang, Binghui Huang, Fuchun Sun and Bin Fang. “Reusing discriminators for encoding: Towards unsupervised image-to-image translation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[30] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford and Xi Chen. “Improved techniques for training gans.” Advances in neural information processing systems 29 (2016).

[31] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler and Sepp Hochreiter. “Gans trained by a two time-scale update rule converge to a local nash equilibrium.” Advances in neural information processing systems 30 (2017).

[32] Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel and Arthur Gretton. “Demystifying mmd gans.” arXiv preprint arXiv:1801.01401 (2018).

[33] Brown, Tom, et al. “Language models are few-shot learners.” Advances in neural information processing systems 33 (2020): 1877-1901.

[34] Zihang Dai, Hanxiao Liu, Quoc V Le and Mingxing Tan. “Coatnet: Marrying convolution and attention for all data sizes.” Advances in Neural Information Processing Systems 34 (2021): 3965-3977.

[35] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. “Mobilenetv2: Inverted residuals and linear bottlenecks.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[36] Liu, Ze, et al. “Swin transformer: Hierarchical vision transformer using shifted windows.” Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[37] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga and Tengyu Ma. “Learning imbalanced datasets with label-distribution-aware margin loss.” Advances in neural information processing systems 32 (2019).

[38] Gu, Hongyu, Zhecheng Yang, and Hong Chen. “Automatic Tongue Image Segmentation Based on Thresholding and an Improved Level Set Model.” 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). IEEE, 2020.

[39] Chao Song, Bin Wang, and Jiatuo Xu. Computer Engineering and “A tongue image classification method based on deep transfer learning.” Science 43.08(2021):1488-1496. (in Chinese)

[40] Jiawei Li et al. “Automatic Classification Framework of Tongue Feature Based on Convolutional Neural Networks.” Micromachines vol. 13,4 501. 24 Mar. 2022, doi:10.3390/mi13040501