Self-supervision Adversarial Learning Network for Liver Lesion Classification

Cong Ma[1], Zhe Liu[1(✉)], Yuqing Song[1], Chengjian Qiu[1], Aihong Yu[1], and Jiawen Zhang[2]

[1] School of Computer Science and Communications Engineering, Jiangsu University, Zhenjiang 212013, China
1000004088@ujs.edu.cn

[2] Huashan Hospital, Department of Radiology, Fudan University, 12 Middle Wulumuqi Road, Shanghai 200040, China

Abstract. The lack of training samples is one of the main factors affecting the development of deep learning methods. Deep learning models often fail to learn useful features and have serious over-fitting problems when lacking of training data. In this work, we exploit two popular unsupervised learning techniques: adversarial learning and self-supervised learning, which is aimed at mine more useful representations and relieve over-fitting problems. Our training scheme is mainly divided into three steps. Firstly, we train a self-supervision network with unsupervised learning to extract obvious features from our liver lesion samples and these features will be transferred to next step. Secondly, we use the final output feature map generated by self-supervision network to train a discriminator by adversarial learning. Finally, the backbone network is trained under the constraint of discriminator and classifier. Our main idea is to train a discriminator with adversarial learning and self-supervised learning. Then, we use the discriminator to constrain the backbone network, which is aimed to reduce the backbone network solution search space. In particular, Different from generating data with GAN, we use GAN to feature adversarial learning for feature augmentation. Our experiments on liver lesion classification in CT show an average accuracy as 92.51% compared with the baseline training scheme, which demonstrates our proposed method can mime useful features and relieve over-fitting problem. It can assist physicians in the early detection and treatment of liver lesions.

Keywords: Self-supervised learning · Adversarial learning · Liver lesions

1 Introduction

The world Health Organization survey shows that the incidence of liver cancer and mortality Ranked 4th and 2nd in the world [1]. Medical image analysis plays a great important role in early diagnosis and treatment of liver tumor. Computed tomography (CT), Magnetic Resonance Imaging (MRI) and Liver Biopsy (LB) are the main methods for clinical analysis and diagnosis of liver tumor. Compared with the latter two methods, CT is often used to assist liver tumor diagnosis because of its robustness, high resolution

X. Sun et al. (Eds.): ICAIS 2021, LNCS 12736, pp. 540–552, 2021.
https://doi.org/10.1007/978-3-030-78609-0_46

and convenience. Focal liver lesion detection and classification are common medical problems, which determines the type of liver lesion. Hepatocellular carcinoma (HCC), hemangiomas (HEM) and metastasis (MET) are common liver lesion types [2].

With the increase in the number of patients, there will be a lot of medical CT images. However, the current number of professional physicians is very small. Traditional manual analysis often requires repeated comparison of CT slices in different periods, which is time-consuming and laborious and relies heavily on the judgment of professional physicians. So that doctors have a heavy workload, which inevitably leads to some misdiagnosis and missed diagnosis. Therefore, it is of great significance to develop computer-aided diagnosis systems which can assist physicians in the early detection and treatment of liver lesions.

In recent years, deep convolutions neural networks (DCNN) have achieved good results in the field of computer vision and also in the medical domain [3]. DCNN can mine the high-level feature representations, which has been proven to be superior to hand-crafted low-level features and mid-level features [4]. In most medical imaging tasks, medical data annotation is usually made by professional physicians and it takes a lot of time and effort, which is the key bottleneck for deep learning methods in medical domain. Researchers attempt to overcome this challenge by using data augmentation schemes, such as translation, rotation, flip and scale. Using such data augmentation schemes to improve the training process of network has become a standard procedure in computer vision tasks [2]. However, little additional information can be gained from small modifications to the images (e.g. the translation of the image a few pixels to the right) [5]. Therefore, how to improve the model performance under limited data remains to be solved.

The conventional approach to deal with this problem is to augment existing data by using data augmentation schemes. In addition to data augmentation, the extracting effective and discriminant features is related to improve model performance [6]. In this research we proposed a self-supervision adversarial learning network to feature extracting within our dataset to improve performance for liver lesion classification. The dataset used includes 4 categories: Hepatocellular carcinoma, cyst, metastasis, hemangioma and healthy liver parenchyma. Our experiments show that the proposed feature augmentation method improves the baseline results in classification accuracy.

Contribution. Our work is the first attempt to combine self-supervised learning [7] with adversarial learning [8] in the medical images, which can help to mine more useful representations for our classification. When we have small medical training samples, our model can avoid complex knowledge transfer and avoid relieve over-fitting. Our experiment show that our model can obviously improve the performance of liver lesion classification in limited medical data compared to baseline (we employ resnet34 [9] as our baseline).

2 Related Work

In recent years, people are focused on the development of computer-aided diagnostic tool, because it can help radiologists to classify different types of lesions [3, 10]. At

the same time, due to deep learning has prominent effects compared with traditional methods on the ImageNet competition [11], which makes deep learning techniques to become more and more popular on medical imaging filed [12].

Recently, a number of approaches have been proposed to address the classification of liver tumors. For example, in [13], the author used synthetic new medical images by employing Generative Adversarial Networks (GAN) and arrived the results about sensitivity and specificity that had increased to 85.7% and 92.4% respectively. Avi Ben-Cohen et al. had synthesis new data by mixing the class specified and unspecified representation of different factors in the training data. They yielded an average improvement of 7.4% in accuracy over the baseline training scheme [14]. Their work showed that synthetic medical images could effectively improve the model performance. However, in the method above the synthesis new data based on existing data to improve the classification performance of liver lesion. K. Yasaka et al. proposed a convolution neural network to achieve the classification based on 1068 lesion CT images and conducted the testing with the models preserved in different processes which yielded an overall accuracy of 84% [3]. Besides, [15] combined patches of lesion region feature with whole-lesion region information, their method obtains an overall accuracy of 87.23% on the dataset which contains 480 CT liver slice images. Although to some extent, the above methods can alleviate few-sample problem, there are still the following problems: (1) The training fluctuates greatly and new synthesized unknown lesions need to be labeled again by the physician. (2) It is difficult for the network to learn the classification characteristics of different lesions.

In [16], the authors employed end-to-end deep learning approach to discriminate liver metastases from colorectal cancer and benign cysts in abdominal CT images of the liver, obtaining an accuracy of 96% and F1-score of 0.92 based on an in-house clinical biobank with 230 liver lesions originating from 63 patients. However, the method only can use pre-trained model on others large dataset, which needs computed intensively. [17] propose a multi-view knowledge-based collaborative (MV-KBC) deep model to separate malignant from benign nodules using limited chest CT data, this method above learns 3-D lung nodule characteristics by decomposing a 3-D nodule into nine fixed views. The method achieved an accuracy of 91.60% for lung nodule classification with an AUC of 95.70% on the benchmark LIDC-IDRI data set. But the method fails to extract fine features.

In order to develop an effective way for liver lesion classification, we focus on extracting features about rich contextual semantic information with multiple approaches in our dataset. We design a self-supervision adversarial learning network. Our main idea is to train a discriminator with adversarial learning and self-supervised learning. Then, we use the discriminator to constrain the backbone network, which is similar to the function of regularization and imposes soft constraints on the backbone network parameters.

3 Methods and Materials

In this section, we provide the details of our system model under small medical training data. We first introduce our learning framework and its training strategy in Sect. 3.1. Then, self-supervised learning is discussed in Sect. 3.2. Finally, we discuss adversarial learning in Sect. 3.3.

3.1 Overall Structure

In this study, we propose a self-supervision adversarial learning network (see Fig. 1). Figure 1 is the overall structure of our model. Our network includes three mainly steps. First of all, we train a SSNet (self-supervised network) (see Fig. 2) using a cross-entropy loss and SSNet will be as a part of step two. The second step we train an adversarial learning model (see Fig. 3) by using adversarial loss. In our adversarial learning model, the self-supervised network as a feature extractor to extract global features and fix the parameters of the self-supervised network. At the same time, the final output feature generated by SSNet and generator are used as true samples and fake samples respectively and generate adversarial training improve the performance of our model in liver lesion classification. The third step we train a generator (see Fig. 4) using a classification loss under the constraint of the discriminator which is fixed parameters in the step two. In order to improve generator liver lesion classification accuracy, we use hyperparameter λ to balance the influence of discriminator to generator (which is used to classify).

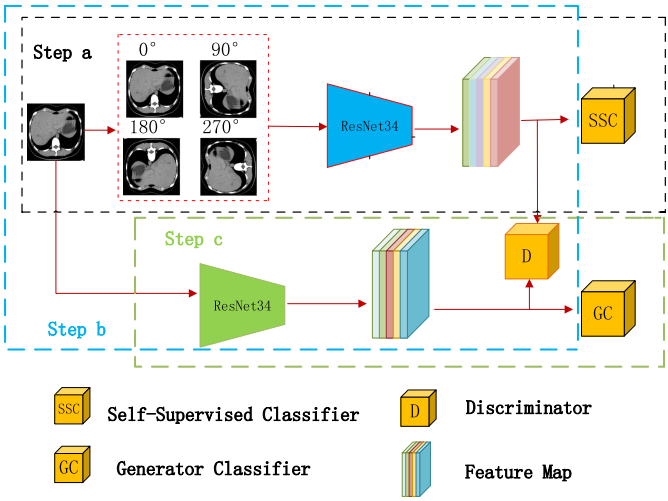

Fig. 1. Our Network structure. In this structure we will perform Self-supervised learning and adversarial learning. Step (a) (b) and (c) perform the self-supervised learning (black dotted box), adversarial learning (blue dotted box) and classification, respectively. (Color figure online)

3.2 Self-supervised Learning

Self-supervised learning (see Fig. 2), which learns by constructing artificial labels given only the input signals, has recently gained considerable attention for learning representations with unlabeled datasets [18], i.e., Gidaris et al. [19] proposed to rotate the image and predict the rotation angle. It has been widely used in the video domain [20], the robotics domain [21] and the image domain [22]. We focused on the medical image domain in this paper. This surrogate task mines useful representations for downstream image classification tasks. There are many other surrogate tasks besides rotating classification. For

example, the network can be trained to solve the context prediction problem, like the relative location of disjoint patches [7]. To predict such transformations, a model should distinguish between what is semantically natural or not. Consequently, it learns high-level semantic representations of inputs [19]. Other surrogate tasks include predicting the unsupervised clustering classes [22], image in-painting [23] and so on.

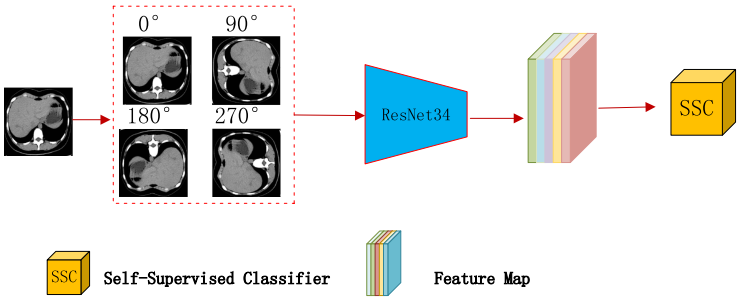

Fig. 2. The process of the self-supervised learning to mine representations. We call the network as SSN (Self-Supervised Network)

3.3 Adversarial Learning

Generative Adversarial Network (GANs) is a class of unsupervised generative models [13]. In [15], GAN is used to generate new samples, which aims to learn the data distribution from a set of samples to generate synthesis liver lesion data drawn from the learned distribution [16]. GAN is often used to augment data to improve the performance of special tasks. Different from generating data with GAN, we employ the thought of the adversarial learning of GAN to augment feature learning [24] (see Fig. 3). The ability of GAN can learn the distribute of data which constraint the feature learning. The key idea of using adversarial learning is to enhance generator to learn useful representation.

In our third training step (see in Fig. 3), we use generate adversarial training to promote the generator for classification task. The feature maps generated by SSNet and generator are viewed as true samples and false samples respectively, and then generate adversarial training for both to improve of generator in liver lesion classification. Using generate adversarial training, we do not need to designed complex cost function and training method to classification task. Our network can learn more better representations from using generate adversarial training. The experiment is shown in Sect. 4.

When we train our model, there are three mainly loss functions including Self-supervised loss, adversarial loss and classifier loss. The adversarial loss is similar to GAN. In GAN, discriminator as D and generator as G, playing the following two-player minimax game, which denotes loss optimization of the generative adversarial model in below [21]:

$$\min_{G}\max_{D} \mathrm{E}_{x\sim \mathrm{P}_{data}}[\log D(x)] + \mathrm{E}_{x\sim \mathrm{P}_{z}}[\log(1 - D(G(z)))] \tag{1}$$

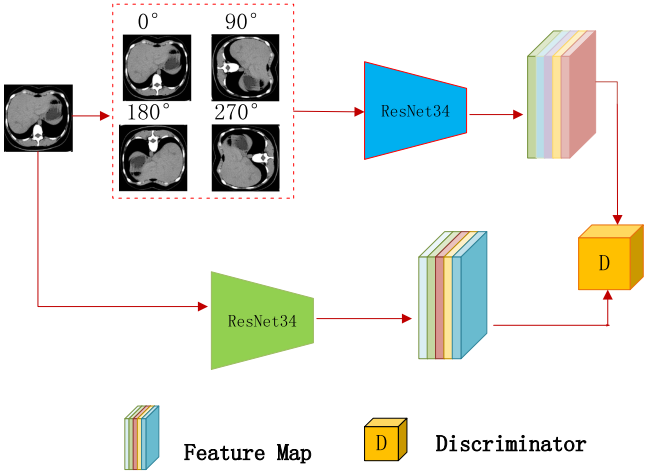

Fig. 3. The process of adversarial learning

where E denotes expectation, x and z are samples drawn from P_{data} and P_z respectively. The discriminator D is trained to maximized D(x) for images x with $x \sim P_{data}$ and to minimize D(G(z)) for images G(z) with $x \sim P_z$. G(z) which denotes the features generated is adopted to fool D during training. Therefore, the generator is trained to maximize D(G(z)).

In our model, when we train the first step, the classifier of SSNet is cross entropy classification loss L_{SSC} is as follows:

$$L_{SSC} = -\frac{1}{N}\sum_{i=1}^{N}[y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log\left(1-\hat{y}^{(i)}\right)] \tag{2}$$

When train the second step, we were inspired by [19], the adversarial loss can be divided into discriminator loss L_D and generator loss L_G.they are as follows:

$$L_D = \max_D \frac{1}{N}\sum_{i=1}^{N}[\log(D(E(x^{(i)}))) + \log(1 - D(G(x^{(i)})))] \tag{3}$$

$$L_G = \max_D \frac{1}{N}\sum_{i=1}^{N}\log(1 - D(G(x^{(i)}))) \tag{4}$$

where N is the batch size, x^i represents the i-th sample (i = 1, 2, …, N). E $(x^{(i)})$ represents the i-th sample output feature maps form SSNet. G $(x^{(i)})$ represents the i-th sample output feature maps from generator.

In the third step (see Fig. 4), the classifier of generator is Cross entropy classification loss L_{GC} (see formula 5)

$$L_{GC} = -\frac{1}{N}\sum_{i=1}^{N}[y^{(i)}\log\hat{y}^{(i)} + (1-y^{(i)})\log\left(1-\hat{y}^{(i)}\right)] \tag{5}$$

In particular, in order to carry out the efficiency liver lesion classification, we use hyperparameter λ to balance the feature earned. So that the third step loss L is as follows:

$$L = \lambda L_D + L_{GC} \tag{6}$$

where λ is set to be negative in order to achieve the desired separation of the representation.

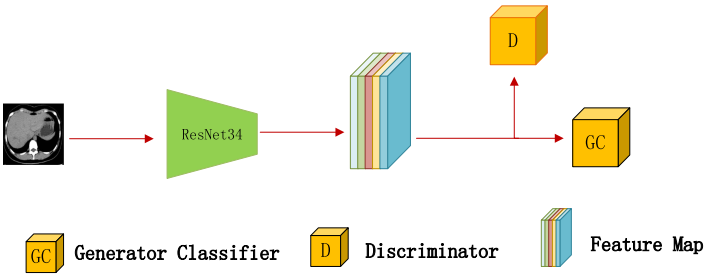

Fig. 4. The process of classification training

4 Experiments and Results

4.1 Dataset and Implementation

In this work, experimental data was obtained from the Affiliated Hospital of Jiangsu University, the CT scans were acquired with a slice collimation of 5–7 mm, a matrix of 512 × 512 pixels, and an inplane resolution of 0.57−0.89. The dataset comprises four types of 430 portal CT scans from 120 patients: 163 metastases (MET), 83 hemangiomas (HEM) and 184 Hepatocellular carcinoma (HCC). In order to increase the diversity of data, 84 samples were taken from the non-marked liver location Image of Healthy, so our final classification dataset have 514 images. Figure 5 shows a set of data samples from the different categories. The lesion sample in the picture is the lesion area taken according to the doctor's mark. The image input to the classification network is a region of interest captured based on the annotations of radiologists. In order to remove irrelevant information about other organs and tissues in the CT scans for liver lesion classification, we cut the image intensity values of all CT scans to the range of [−100, 400] HU [24]. After all CT scan HU values were truncated, we normalized all slice intensities into the range [0, 1] with min-max normalization.

We set batch size is 64 and the learning rate is 0.001 for 50 epochs. In each epoch, our training is divided into three steps, when train the first step, we perform 20 epochs, after that we perform adversarial learning with 25 epochs. In the third step, we train classification model 10 epochs. The input to our classification system is 64 × 64 image from the region of interests by random cropping. All training processes were performed using a NVIDIA GeForce GTX 1080 Ti GPU.

4.2 Evaluation

We use 5-fold cross validation to evaluate the classification performance with and without our proposed method, our training perform on the dataset which is included in Sect. 4.1. Healthy tissues are used to enhance the diversity of training data and make the model have the judgment ability on lesions and non-lesions. The results are presented in Table 1. By using this training strategy, we achieved an improvement of 10.2% on average accuracy.

Table 1. Results with and without K-fold cross-validation were compared in the final epoch.

Fold	Baseline [ACC.%]	Proposed [ACC.%]
1	83.68	88.81
2	79.53	90.09
3	82.16	94.60
4	83.32	92.46
5	81.23	95.44
Avg	82.01 ± 2.2	92.28 ± 3.1

Classification confusion matrix with and without our proposed strategy are presented (see Fig. 5 and Fig. 6). From the results, we can see that our proposed method was able to improve the overall classification accuracy along all classes. However, there are some mistakes between MET, HEM, Normal. Figure 7 shows the comparison in baseline and our proposed method with ROC curves. Our proposed method improves AUC along all classes.

In order to explore the impact of different strategies on liver lesion classification, we compared baseline, baseline with adversarial learning, baseline with self-supervision and our proposed method. Table 2 has showed the performance comparison of models with different strategies. We can find self-supervision can achieve better result than using the adversarial learning.

However, there was a serious over-fitting problem when we trained baseline with self-supervision. In order to alleviate over-fitting problem in our dataset, we use generate adversarial training to promote model self-supervision to relieve over-fitting problem. We did an experiment to compare feature learning based on adversarial constraint. It can be found that the baseline with feature adversarial constraint and self-supervision has better performance than baseline with self-supervision.

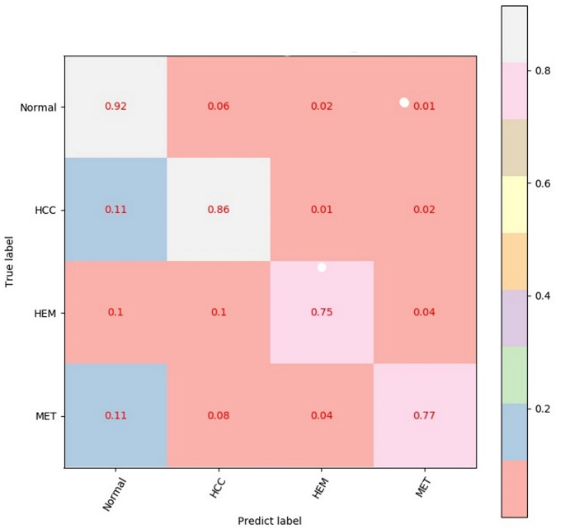

Fig. 5. Mixed matrix for the baseline

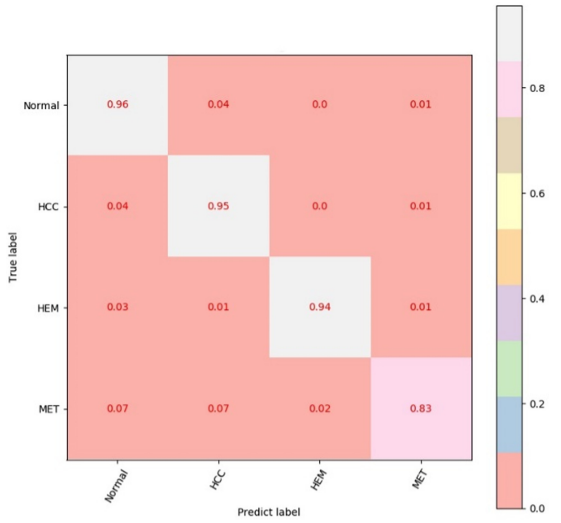

Fig. 6. Mixed matrix for our proposed method

Previous studies have proven that deep learning based on methods are superior to traditional methods, people more tend to use deep learning for liver lesions classification [3, 4, 15]. Due to the limited number of openly available liver lesion dataset with confirmed presence of malignant tumors, direct comparative with other published results is difficult. However, we comparative results reported in this study show the improvement with regards to state of art methods, including classic network. The results are presented

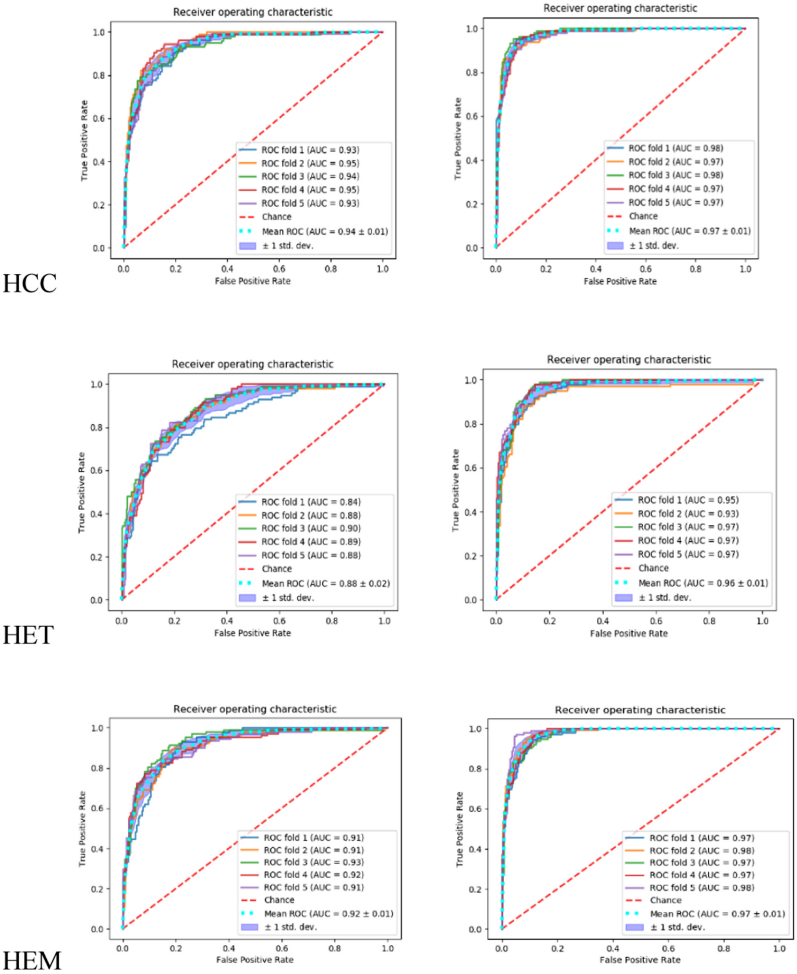

Fig. 7. Comparison in baseline and our proposed method with ROC curves. Left column show the results of baseline and Right column show the results of our method. Our method improves AUC along all classes.

in Table 3. It shows the proposed method, which surpasses the previous classic network model.

Table 2. Classification accuracy (%) of Baseline with self-supervision (B + SS), Baseline with adversarial learning (B + AL), and our proposed which the best accuracy.

Model	Accuracy	Recall	Precision	F1-score
Baseline	83.2 ± 1.3	81.9 ± 2.2	80.4 ± 0.8	79.6 ± 2.2
B+AL	86.6 ± 2.5	85.6 ± 2.7	85.3 ± 1.7	84.3 ± 1.9
B+SS	89.5 ± 3.1	86.3 ± 2.7	88.4 ± 1.2	86.7 ± 3.6
Proposed	92.5 ± 2.8	91.2 ± 2.1	90.5 ± 0.6	89.5 ± 1.9

Table 3. Proposed method comparison with mainstream classification models

Model	Accuracy	Recall	Precision	F1-score	AUC
ResNet18 [9]	79.2 ± 0.6	75.9 ± 1.5	78.3 ± 1.8	74.5 ± 2.6	93.6 ± 2.2
ResNet101 [9]	81.6 ± 1.1	79.2 ± 2.8	80.6 ± 3.5	80.6 ± 3.5	95.1 ± 3.1
SENet34 [25]	78.9 ± 0.9	75.4 ± 1.2	74.9 ± 2.1	70.4 ± 5.7	93.5 ± 1.5
SENet50 [25]	80.1 ± 1.8	76.7 ± 3.5	76.1 ± 1.4	77.4 ± 2.2	94.1 ± 1.75
SE_ResNet34 [26]	83.5 ± 3.3	80.3 ± 3.6	81.6 ± 1.8	79.2 ± 3.5	96.7 ± 1.9
Proposed	92.5 ± 0.8	91.1 ± 2.1	90.2 ± 0.6	89.5 ± 1.9	98.0 ± 0.1

5 Conclusions

In this study, we proposed a self-supervision adversarial learning network to classify different liver lesions in CT images. Using self-supervised learning and adversarial learning scheme, we can mine the useful feature representations in limited training data for liver lesion classification. Our methods far exceed the previous proposed approaches accuracy and AUC. Therefore, our framework can provide a screening tool for early detection of malignant lesions.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (61976106, 61572239), Zhenjiang Key Deprogram "Fire Early Warning Technology Based on Multimodal Data Analysis" (SH2020011) Jiangsu Emergency Management Science and Technology Project "Research on Very Early Warning of Fire Based on Multi-modal Data Analysis and Multi-Intelligent Body Technology" (YJGL-TG-2020-8).

References

1. Bray, F., Ferlay, J., Soerjomataram, I., et al.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: Cancer J. Clin. **68**(6), 394–424 (2018)
2. Greenspan, H., Van Ginneken, B., Summers, R.M.: Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. IEEE Trans. Med. Imaging **35**(5), 1153–1159 (2016)

3. Yasaka, K., Akai, H., Abe, O., et al.: Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology **286**(3), 887–896 (2018)
4. Setio, A.A.A., Ciompi, F., Litjens, G., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging **35**(5), 1160–1169 (2016)
5. Frid-Adar, M., Diamant, I., Klang, E., et al.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing **321**, 321–331 (2018)
6. Luo, L., Xiong, Y., Liu, Y., et al.: Adaptive gradient methods with dynamic bound of learning rate. arXiv preprint arXiv:1902.09843 (2019)
7. Doersch, C., Gupta, A., Efros, A.A.: Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1422–1430 (2015)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
9. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Adcock, A., Rubin, D., Carlsson, G.: Classification of hepatic lesions using the matching metric. Comput. Vis. Image Underst. **121**, 36–42 (2014)
11. Deng, J., Dong, W., Socher, R., et al. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255 (2009)
12. Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
13. Frid-Adar, M., Klang, E., Amitai, M., et al.: Synthetic data augmentation using GAN for improved liver lesion classification. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 289–293. IEEE (2018)
14. Ben-Cohen, A., Mechrez, R., Yedidia, N., et al.: Improving CNN training using disentanglement for liver lesion classification in CT. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 886–889. IEEE (2019)
15. Liang, D., et al.: Residual convolutional neural networks with global and local pathways for classification of focal liver lesions. In: Geng, X., Kang, B.-H. (eds.) PRICAI 2018: Trends in Artificial Intelligence: 15th Pacific Rim International Conference on Artificial Intelligence, Nanjing, China, August 28–31, 2018, Proceedings, Part I, pp. 617–628. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97304-3_47
16. Romero, F.P., Diler, A., Bisson-Gregoire, G., et al.: End-to-end discriminative deep network for liver lesion classification. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1243–1246. IEEE (2019)
17. Xie, Y., Xia, Y., Zhang, J., et al.: Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. IEEE Trans. Med. Imaging **38**(4), 991–1004 (2018)
18. Lee, H., Hwang, S.J., Shin, J.: Self-supervised label augmentation via input transformations. In: International Conference on Machine Learning, PMLR, pp. 5714–5724 (2020)
19. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728 (2018)
20. Lee, H.Y., Huang, J.B., Singh, M., et al.: Unsupervised representation learning by sorting sequences. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 667–676 (2017)
21. Jang, E., Devin, C., Vanhoucke, V., et al.: Grasp2vec: learning object representations from self-supervised grasping. arXiv preprint arXiv:1811.06964 (2018)

22. Caron, M., Bojanowski, P., Joulin, A., et al.: Deep clustering for unsupervised learning of visual features. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 132–149 (2018)
23. Pathak, D., Krahenbuhl, P., Donahue, J., et al.: Context encoders: feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536–2544 (2016)
24. Chen, P., Song, Y., Yuan, D., et al.: Feature fusion adversarial learning network for liver lesion classification. In: Proceedings of the ACM Multimedia Asia, pp. 1–7 (2019)
25. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
26. Li, X., Wang, W., Hu, X., et al.: Selective kernel networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 510–519 (2019)