SCCNN: A Diagnosis Method for Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma Based on Siamese Cross Contrast Neural Network

QIYUAN WANG1, ZHONGMIN WANG2, YU SUN1, XIN ZHANG1,2, WEIFENG LI1, YUN GE1, XIAOLIN HUANG1, YUN LIU2, AND YING CHEN1

1School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
2Department of Information, First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China

Corresponding authors: Yun Ge (geyun@nju.edu.cn) and Ying Chen (yingchen@nju.edu.cn)

This work was supported in part by the Social Development Program of Primary Research and Development Plan in Jiangsu Province under Grant BE2016733 and Grant BE2017679, in part by the Jiangsu Provincial Basic Research Programme (Natural Science Fund) under Grant BK20190309, and in part by the National Natural Science Foundation of China under Grant 81671777.

ABSTRACT This paper proposes a novel siamese cross contrast neural network (SCCNN) to classify the hepatocellular carcinoma (HCC) and the intrahepatic cholangiocarcinoma (ICC) on computed tomography (CT) images. This method is inspired from cross contrast neural networks (CCNN) which is based on tailored CNN and information based similarity (IBS) theory. A new IBS-based measurement named as discriminative IBS (DisIBS) is designed for SCCNN. SCCNN is composed of two main parts including siamese feature extractors with DisIBS operator and MLP classifiers. Siamese networks extract features with DisIBS calculated by DisIBS operator as metric at the top. MLP classifiers are connected with but gradient-stop to feature extractors deriving classification results. We assign different loss functions with different parts to make better practice, specially DisIBS-based loss for feature extractors and softmax-based for MLP classifiers. SCCNN preserves the advantages of CCNN that can fit the insufficient medical images and small lesions. Furthermore, it extends CCNN with the siamese mechanism and gradient-stop MLP classifiers to accept the random inputs and predict like traditional CNN. To present the effectiveness of SCCNN empirically, we apply this method on a 234-person (157/77 for train/test) dataset and achieve better results than other classic CNN and CCNN methods. We try different base models of siamese structures and display prediction accuracy in two levels (slice/patient). The highest slice/patient accuracy which we have achieved on three-categories classification (HCC/ICC/Normal) is 90.22%/94.92% and the accuracy rises to 94.17%/97.44% on binary classification (HCC/ICC).

INDEX TERMS Siamese networks, multiple loss, cross contrast neural networks, hepatocellular carcinoma, intrahepatic cholangiocarcinoma.

I. INTRODUCTION

Over the past few years, the incidence and mortality of liver cancer have increased rapidly. Liver cancer has become one of the most common diagnosed cancers and held the second place in terms of deaths for males [1]. This disease in China is even more serious, occupying over 50% of new cases and deaths worldwide [2]. Primary liver cancer includes hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma as well as other rare types [3]–[6]. Accounting for over three quarters (75%–85%) of liver cancer cases, HCC is the most common primary liver cancer, followed by ICC (10%–15% of cases). HCC is different from ICC in biological behaviors which greatly affect the subsequent treatment plan and prognosis [7]. It is critical for the clinical diagnosis and treatment of liver cancer to classify HCC/ICC accurately.

In the context of histological HCC/ICC sub-classes, each with distinct molecular patterns and prognostic impacts, liver biopsy is still the gold-standard method to distinguish HCC/ICC [8]. However, liver biopsy is an invasive method and suffered in the sampling error of small biopsies [9]. The rapid development of imaging techniques provides a new way to implement a noninvasive diagnosis for liver cancer. However, there have been relatively few researches lying in
distinguishing HCC/ICC with images until now. The existing studies try to make breakthroughs on ultrasonography, computed tomography (CT), and magnetic resonance imaging (MRI) images.

Tsunematsu et al. [10] used the Chi square test as well as the non-parametric Mann-Whitney U test which were common feature-selection methods to analysis the preselected features, 14 clinical characteristics and some CT imaging features. The logistic regression analysis was applied to find the best features to distinguish between ICC and HCC. They finally selected the most powerful feature, the presence of an intratumoral artery, and achieved 78.9% accuracy.

They finally selected the most powerful feature, the presence of an intratumoral artery, and achieved 78.9% accuracy on dynamic CT. Yi Wei et al. [11] differed ICC and HCC by using the selected Intravoxel incoherent motion (IVIM) and diffusion weighted imaging (DWI) parameters on MRI images, such as true ADC derived from bi-exponential model. An iterative Fuzzy C-Mean method was proposed to distinguish a focal liver lesion from a liver ultrasound image [12]. Dwidar et al. [13] evaluated the use of liver and spleen stiffness measured by ultrasound shear wave elastography (LSSWE, SS-SWE) in CLD (chronic liver disease) diagnosis.

Some machine learning methods were also applied in related works. Virmani et al. [14] proposed a SVM-based/support vector machine method which uses wavelet transform-based texture features on liver ultrasound images, increasing the classification accuracy to 88.8%. Mougiakakou et al. [15] proposed a computer-aided diagnosis architecture, based on two ensembles of classifiers for the classification of liver tissue. The best performance of the architecture achieved a mean classification accuracy 84.96%. Mitrea et al. [16] and Xian [17] proposed new methods for evaluating the texture features, based on the gray level co-occurrence matrix. An identification method of malignant and benign liver tumors based on fuzzy support vector machine was proposed by Xian et al. Ashour et al. [18] applied a subspace/discriminant ensemble classifier on liver fibrosis level staging.

The achievements of deep learning methods in computer vision and pattern recognition inspire researchers to make a new attempt in liver cancer classification. Wang et al. [19] proposed a CNN-based classifier whose accuracy results on liver microscopic images for the normal class, granuloma-fibrosis1, and granuloma-fibrosis2 were 92.5%, 76.67%, and 79.17%, respectively. Midya et al. [20] used a pre-trained network by retraining the final layers to derive the best accuracy 69.7% and area under the receiver operating characteristics curve 0.72 on a 223-patient dataset.

All these methods above show that noninvasive liver cancer diagnosis on images is possible and researchers have made some achievements in related works. However, most of these methods only extract dominant features in tumor area, such as histogram, first-degree, second-degree gray level statistics of tumor area and more complex morphological features derived from wavelet transform, co-occurrence matrix and GLCM. They are sensitive to the size of tumor area and perform poorly in the case of small lesions. Meanwhile, it’s difficult for them to fit the insufficient medical datasets problem and easy to suffer in overfitting. Huang et al. [21] proposed a cross contrast neural networks which combine the powerful capability of CNN with the IBS statistic analysis method [22] to overcome the difficulties above and successfully applied it on staging liver fibrosis. The effectiveness of CCNN inspires us to propose the SCCNN. Compared with CCNN, our method preserves the advantages that can fit the insufficient medical datasets and be insensitive to small lesions. Furthermore, owing to the siamese mechanism and gradient-stop MLP classifiers, SCCNN can accept random inputs and predict like traditional CNN in the test process.

Contributions of this study can be summarized as follows:

1. In this study, we extend CCNN with the siamese mechanism and gradient-stop MLP networks. These strategies introduce two primary advantages to SCCNN. One is that SCCNN can accept the random inputs. The other is that one test example can be fed into any one of the siamese structures and derive the direct classification result in test process. Each of the siamese structures, especially one feature extractor together with corresponding MLP classifier, can be used like traditional CNN.

2. We assign different loss functions with different parts of SCCNN to make better practice. Feature extractors, siamese convolutional neural networks, are only optimized by a DisIBS-based loss to extract more discriminative features. A common softmax-based loss pushes the gradient-stop MLP classifiers mapping the extracted features to accurate classification results. Most noteworthy is that the softmax-based loss can’t influence the former because of gradient-stop.

3. We present empirically a noninvasive method which can diagnose HCC, ICC and normal accurately on CT images. SCCNN improve the results by 11.6%/10.17% (slice/patient) on binary classification and 8.39%/8.24% (slice/patient) on three-categories classification compared with other methods in this task.

With the empirical study on a real clinical dataset, the superior results demonstrate the effectiveness of our proposed framework. This paper is organized as follows. Section II presents some related works including the siamese networks and CCNN. Section III describes the dataset materials and the proposed SCCNN. Section IV and V show the experiment details, specific results and related visualizations. Section VI and VII conclude the attributes of our framework and extend some directions to research further.

II. RELATED WORKS

A. SIAMESE NETWORKS

Siamese networks were first introduced in the early 1990s by Bromley et al. [23] to solve signature verification as an image matching problem. A siamese neural network is composed of parameters-tied twin networks which accept pair inputs but are joined by an energy function at the top. Some specific functions are designed to use the highest layer...
features to calculate metrics between pair inputs. The capabilities of different siamese-based methods depend on four main points: the network architecture, the training set selection strategy, the metric function, and the training algorithm. Because siamese networks not only use the information of the image itself but also capture information from the contrast of image pairs, they have shown powerful potential in the medical image diagnosis where the number of examples is small and the inner classes have high variances. Amin-Naji et al. [24] apply siamese convolutional neural network to diagnosis Alzheimer’s Disease (AD) on 235 subjects from OASIS dataset and achieve 98.72% accuracy. Furthermore, Siamese CNNs have been used successfully in various tasks, such as person re-identification [25], speaker verification [26] and face identification [27].

### B. CROSS CONTRAST NEURAL NETWORKS

Cross contrast neural networks were first proposed by Huang et al. [21] to stage liver fibrosis. This method consists of two main parts. The first part applies convolutional neural networks to extract features and gets the cross probability maps for utilizing the implicit contrast information among the inputs. The second part measures the similarity between two maps using the modified information based similarity (IBS) theory [22]. IBS theory is a statistical method which quantifies similarity between symbols based on the order of occurrence of a particular pattern and has been proved effective in many areas [28], [29]. The main assumption is that the content of information in any symbol sequence is mainly determined by the frequency of reuse of its essential elements. CCNN combines the advantages of statistical analysis and convolutional neural networks, fitting the insufficient medical images and being insensitive to small lesions. They applied the method on a 34-person dataset, achieving the highest accuracy 98.3% on binary classification and 84.4% on five-categories classification. You can find more information about the methods and algorithms in the original paper. The super-category classification. You can find more information about the methods and algorithms in the original paper.

### III. MATERIALS AND METHODS

#### A. MATERIALS

The observational studies conducted in this article have been approved by the Jiangsu Provincial Hospital ethics committee, and written informed consent was obtained from each patient. The patients involved in this study were diagnosed from January 2014 to September 2017. All these patients underwent partial hepatectomy and had been confirmed by liver biopsy. Pathological analysis of biopsy sections of liver tissue is the diagnosis result given by a pathologist who has 10 years experience in hepatobiliary surgery. 82 HCC cases, 73 ICC cases together with 79 normal cases without liver diseases form the dataset. Table. 1 shows the specific statistics of patients. In this study, we use the dynamic CT slice images before hepatectomy for liver diagnosis.

#### B. METHODS

1) **STRUCTURE**

From the structural capability perspective, SCCNN is composed of two main parts including siamese feature extractors with DisIBS operator and MLP classifiers. Feature extractors map the image space to feature space and extract discriminative features. To take the advantages of siamese mechanism, SCCNN exploit twin parameters-tied parallel CNNs as feature extractors. Instead of some value sensitive measurements such as Euclidean distance/Manhattan distance, SCCNN take the DisIBS which is a new designed IBS-based measurement as the metric function of siamese structure at the top. DisIBS operator accepts a pair of feature probability vectors transformed from the feature extractors to calculate the DisIBS between the pair. The details of DisIBS and feature probability vectors will be presented in next part. Each MLP classifier takes one of the paired feature probability vectors to output the classification result. All MLP classifiers are connected with but gradient-stop to the corresponding feature extractors. What noteworthy is that MLP classifiers are parameters-tied. The entire structure of SCCNN is shown in Fig. 1.

The siamese structures are distributed vertical in Fig. 1. As illustrated in this figure, the green, red, blue lines represent respectively feature extractors, DisIBS operator, and MLP classifiers process. For training, we pair each train example with the other randomly to form the homologous or heterologous pair in turn. For each pipeline, the input is random and the connected MLP classifier outputs the classification result. DisIBS operator calculates the DisIBS value between pair inputs. The entire structure of SCCNN is optimized by multiple loss functions for different parts which will be discussed in the third part of this section. For testing, we can apply any one of the siamese structures, specifically one feature extractor together with the corresponding MLP classifier, to accept one example and output the prediction. Fig.2 shows the test process.

### Table 1. Statistics of patients. A total of 234 people were eventually used in the experimental study, with an average age of 59.2±10 years and an age range of 32-82 years. People in the training set and in the test set do not overlap.

| Type    | Age   | Subject Number  | Number of slices |
|---------|-------|-----------------|------------------|
| HCC     | 61.2±10.3 / 59.2±5.1 | (50,11) / (14,7) | 394 / 107        |
| ICC     | 57.0±8.2 / 58.0±6.8 | (30,25) / (10,8) | 359 / 99         |
| Normal  | 58.3±7.6 / 60.4±7.2 | (35,16) / (18,10)| 298 / 101        |

Q. Wang et al.: SCCNN: A Diagnosis Method for HCC and Intrahepatic Cholangiocarcinoma
2) DisIBS

Inspired from the ModIBS of CCNN [21], we propose a new IBS-based measurement called DisIBS as the metric function of siamese structure at the top. Allowing for some assumptions of IBS, DisIBS cares more about the distribution and rank correlation rather than the magnitude of the features. We calculate DisIBS based on feature probability vectors and rank vectors transformed from the former. They represent the distribution and rank correlation of features respectively. Feature probability vectors $\hat{p}_n$ formulated as Equation (1) are calculated from feature maps and consistent with that of ModIBS. The value of rank vectors in each location is the order of the value of feature probability vectors in the corresponding location.

$$\hat{p}_n = [p_1 \ldots p_i \ldots p_D]_n \quad n = [1, M]$$  

(1)
We change ModIBS and define DisIBS as follow:

\[ p_i = g(f_i) = \frac{\sum_j \text{softsign}(\max(f_{ij}, 0))}{\sum_i \sum_j \text{softsign}(\max(f_{ij}, 0))} \]

\[ R_{x,i} = \text{Rank}(p_{x,i}) \]

\[ \text{margin} = \sqrt{\frac{1}{\ln 0.5} - 1} \]

\[ \text{NorIBS} = \frac{1}{D} \sum_i |R_{x,i} - R_{y,i}| F_i \]

\[ F_i = -E(p_{x,i}) - E(p_{y,i}) \]

\[ \text{RevIBS} = \text{NorIBS}(\text{random}(\tilde{p}_i)) \]

As is Equation (5) shown, DisIBS is positive correlation with NorIBS and its value boundary is closer to 0 or 1. For homologous pairs, the NorIBS is close to 0 and the RevIBS is much larger than the NorIBS, which makes the DisIBS tends to be 0. For heterologous pairs, the RevIBS is far smaller than NorIBS value, which makes DisIBS close to 1. margin is a value related to \( \alpha \) to make training process more stable and adjusts DisIBS close to 0.5 when NorIBS is equal to RevIBS. When compared with ModIBS, it’s similar for DisIBS to push NorIBS closer to 0 for homologous pairs. Different situations occur for heterologous pairs that DisIBS makes NorIBS as large as possible instead of moving closer to RevIBS like ModIBS. It’s helpful for SCCNN to extract more discriminative features. The differences between DisIBS and ModIBS in feature distribution after training are visualized in Fig 3. In this paper, we use a probability distribution map to describe the relation of two feature distributions. The horizontal axis and vertical axis of probability distribution maps are respectively the two probability distribution vectors (Equation (2)) of different images. It’s apparent that DisIBS makes the feature distributions of heterologous pairs move to different directions while ModIBS tends to make the feature distributions irrelevant.

3) MULTIPLE LOSS

In this study, we assign different parts with different loss functions to make better practice. Specially, the DisIBS-based loss is applied to optimize feature extractors extracting more discriminative features while softmax-based losses are used to push MLP classifiers mapping the extracted features to accurate classification results. DisIBS operator is non-parameter module. The parameters of feature extractors are only influenced by DisIBS-based loss because MLP classifiers are connected with but gradient-stop to the corresponding feature extractors. It means that the features extractors are found based on DisIBS theory. Consulting the loss of CCNN, we present a cross entropy formulation loss based on DisIBS.
For testing, SCCNN can accept one image and predict as follow to make train process continuous. A joint loss composed of weighted various losses is defined and computes the fiers accepts one of the paired vectors to make predictions. Feature extractors deriving the extracted feature vectors. DisIBS operator outputs the DisIBS value between the paired label represents whether the pair originates from the same class or not. 0 means the homologous inputs and 1 means heterologous. For each MLP classifier, the classification loss ($\text{Loss}_{\text{softmax}}$) is applied to optimize the parameters. logit represents the predictions of MLP classifier and label$_{\text{class}}$ means the corresponding true class.

\[
\text{Loss}_{\text{softmax}} = (1 - \text{label}_{\text{class}}) \cdot \log(1 - \text{logit}) + \text{label}_{\text{class}} \cdot \log(\text{logit})
\]

For training SCCNN, paired inputs first go through feature extractors deriving the extracted feature vectors. DisIBS operator outputs the DisIBS value between the paired feature vectors and calculates $\text{Loss}_{\text{DisIBS}}$. Each MLP classifier accepts one of the paired vectors to make predictions and computes the $\text{Loss}_{\text{softmax}}$ coupled with the true class. A joint loss composed of weighted various losses is defined as follow to make train process continuous. $\text{Loss}_{\text{softmax}1}$ and $\text{Loss}_{\text{softmax}2}$ separately correspond to two MLP classifiers. $w_{\text{DisIBS}}$, $w_{\text{softmax}1}$, and $w_{\text{softmax}2}$ are weights of different losses. For testing, SCCNN can accept one image and predict like traditional CNN by one feature extractor together with the corresponding MLP classifier.

\[
\text{Loss} = w_{\text{DisIBS}} \cdot \text{Loss}_{\text{DisIBS}} + w_{\text{softmax}1} \cdot \text{Loss}_{\text{softmax}1} + w_{\text{softmax}2} \cdot \text{Loss}_{\text{softmax}2}
\]

### IV. EXPERIMENTS AND RESULTS

#### A. DATASETS AND DATA PREPROCESSING

1) DATASETS
As introduced in Materials part, we use a dataset which is composed of 82 HCC cases, 73 ICC cases and 79 normal cases as controls. With a common-used split ratio about 2.5:1, we use 167 patients for training and 67 patients for testing. For each patient, the results of diagnosis have been confirmed by liver biopsy and all imaging resources have been collected. In this study, we choose the dynamic CT slice images before hepatectomy to form the dataset. For each patient, we select 4-7 slices which are full-shape and have appropriate differences. All slices of the same patient are labeled the same. Finally, there are 394/107 (train/test) HCC slices, 359/99 (train/test) ICC slices, and 298/101 (train/test) control slices. Train and test sets are split by patients to make sure not to overlap.

2) DATA PREPROCESSING
We perform the image segmentation to avoid the interference of extrahepatic tissue since the main task of this study is liver cancer classification. In fact, it’s a challenging task for even doctors to diagnose HCC, ICC and Normal through the liver entities on dynamic CT images. So, in early process of this study, we pay more attention to the main task and take the easy-implemented segmentation method, semi-automated snake algorithm [30]. This method needs manual hand operation but few annotations(templates) to fit accuracy requirements. A template made up of several initial points is manually provided by an experienced radiologist to control the elastic deformation. All slices contain a clear and relatively complete liver entity. For clinical application, an automated segmentation method may fit SCCNN better. Other state of art segmentation methods will be tried in the following study.

After segmentation, we resize the slices as 224*224 or 299*299 because of the fixed input dimension of base model. Considering the potential bias and the raising likelihood of overfitting introduced by the multiple slices of one instance, we apply the data augmentation strategy in this paper. In details, we choose to randomly flip in horizontal, randomly flip in vertical, randomly adjust colorJitter (all ranges = $-0.1 \sim +0.1$) and randomly rotate from $-45^\circ$ to $45^\circ$, as shown in Fig. 4. We also experiment the magnification/shrink and crop operations in data augmentation. However, this strategy can’t increase and even decrease the performance of SCCNN in this task. These operations may not fit this dataset. Instead of increasing fixed times learning data, we do data augmentation in parallel with training process. The advantage of this strategy is that the number of augmentation dataset is increasing with the number of epochs. This can form a larger augmentation dataset. All augmentation operations are only applied in train set. In addition, all slices need to make a reduce-mean operation before fed into SCCNN.

---

**FIGURE 4.** Data preprocessing. The CT image is chosen from a 65-year-old male HCC patient. The data preprocessing for medical CT images includes image segmentation and data augmentation.
TABLE 2. Effectiveness for different base models of SCCNN. We experiment SCCNN with vgg19, inception_v3, densenet202, resnext101, senet_resnet152, resnet152 as base model and evaluate them in binary and three-categories classification for all metrics.

| SCCNN base model | HCC vs. ICC | HCC vs. ICC vs.Normal |
|------------------|------------|----------------------|
|                  | slice      | patient              | slice      | patient              |
| vgg19            | 0.859223   | 0.881667             | 0.863192   | 0.881356             |
| inception_v3     | 0.834176   | 0.845258             | 0.866667   | 0.869312             |
| densenet201      | 0.883495   | 0.932232             | 0.859935   | 0.874576             |
| resnext101       | 0.893204   | 0.871795             | 0.872964   | 0.896274             |
| senet_resnet152  | 0.902913   | 0.902432             | 0.872964   | 0.898305             |
| resnet152        | 0.941748   | 0.974359             | 0.90228    | 0.949152             |

slices which are classified correctly. The prediction of certain patient is the category with the largest number of slices. We vote the slice predictions and set the category with the largest number of slice predictions as patient prediction. The patient level accuracy is calculated by the patient prediction.

C. RESULTS OF SCCNN

SCCNN is flexible and can be easily adjusted by exploring various base models. To make best practice, we experiment the vgg19 [31], inception_v3 [32], resnet152 [33], resnext101 [34], densenet201 [35] and senet_resnet152 [36] as base model of SCCNN for binary (HCC vs. ICC) and three-categories (HCC vs. ICC vs. Normal) classification empirically. We calculate the prediction accuracy in slice and patient levels respectively. In this study, we implement all experiments with PyTorch 1.10 on Tesla P100. We use SGD optimizer with initializing 0.01 and decreasing 0.8 every 5 epochs to get better results. To prevent from overfitting, we apply L2 regularization on all parameters with weighted coefficient 0.004 for binary and 0.0004 for three-categories. We randomly split train/test sets for many times and select the mean accuracy as the final result. The results are demonstrated in Table 2.

As can be seen from the Table 2, the resnet152-based SCCNN holds the first place for both binary and three-categories classification in all metrics. Closer inspection of the HCC vs. ICC column shows that the highest accuracy of the resnet152-based SCCNN reaches 0.941748/0.974359 (slice/patient) and 3.89%/7.19% higher than the second, senet_resnet152-based. Inception_v3-based SCCNN perform worse than others, which is 10.76%/12.91% lower than the best structure. When analyzing the data in HCC vs. ICC vs. Normal column, the best performance of resnet152-based SCCNN achieves 0.90228/0.949152 (slice/patient). The performance of senet_resnet152-based remains steadily second and just a little higher than resnext101-based. The other three models reach similar accuracy. Comparing the same locations between two columns, the values in HCC vs.ICC vs. Normal are lower than those in the other. In each column, patient level accuracy is commonly better than that of slice level for different base models. We also apply the resnet152-based model to configure some other binary classifications, HCC vs. Normal, ICC vs. Normal and HCC+ICC vs. Normal to show generality. The accuracy reaches respectively 0.932692, 0.970000 and 0.947883 in slice level.

D. COMPARISONS WITH OTHER NETWORKS

To verify the effectiveness of our model, we compare our model SCCNN with some classic effective CNNs (including vgg19_bn, resnet152, and inception_v3) and CCNN. To exclude the possible influence of the number of parameters in accuracy, we add the classic CNN with the similar MLP classifiers to SCCNN for comparing. We use a 4-hidden layers MLP classifiers and the hidden neurons are respectively 1024, 256, 256 and 128. The input dimension of MLP classifiers is based on the former CNN structure. The binary and three-categories performance in slice/patient levels are presented in Table 3. As can be seen in this table, SCCNN performs better than others in this task. The highest accuracy of SCCNN for three-categories classification reaches 0.90228/0.949152(slice/patient) and the results rise to 0.941748/0.974359(slice/patient) for binary classification. Among all the methods, CCNN takes the second place and just next to SCCNN. The binary accuracy and three-categories accuracy of SCCNN are respectively 11.6%/10.17% higher and 8.39%/8.24% higher than CCNN. Further analysis on the performance of classic CNN methods including vgg19_bn, resnet152 and inception_v3 shows that resnet152 achieve better results than the other two. Compared
with SCCNN, the accuracy of resnet152 is 13.74%/10.49% lower for binary classification and 8.19%/9.37% lower for three-categories.

Compared with other CNN methods and CCNN, the effectiveness of SCCNN may be on account of three reasons, the siamese structure, DisIBS metric and MLP classifiers. We make ablation studies to study the influence of each factor. siamese-vgg19_bn, siamese-resnet152, siamese-inception_v3 are structures which replace the DisIBS with a Manhattan-distance-based function for SCCNN. SCCNN(without MLP classifiers) are variants which delete the MLP classifiers and use the only DisIBS to make a prediction like CCNN. We compare resnet152, vgg19_bn, inception_v3 respectively with siamese-resnet152, siamese-vgg19_bn, siamese-inception_v3 to study the influence of the siamese strategy. The results illustrate that the siamese strategy alone generally improve the performance but just about 2%. We compare SCCNN with SCCNN (without the MLP classifiers) to study the influence of MLP classifiers. There is a tiny rise in accuracy with MLP classifiers. We compare SCCNN with siamese-resnet152 to study the effectiveness of DisIBS metric. With DisIBS, SCCNN increases the binary accuracy and three-categories accuracy of siamese-resnet152 respectively by 13.1%/9.7% and 8.3%/7.2%. The results show that DisIBS metric is most powerful reason which makes SCCNN reach the high performance in this task and the other two strategies also help increase the accuracy. Nevertheless, the main advantage of MLP classifiers is making SCCNN predict like traditional CNN other than improving the performance.

DisIBS metric in concert with the siamese structure mainly make SCCNN perform better in this task. In one hand, DisIBS is used as the priori knowledge to guide the siamese structure to extract more discriminative features in this paper. In other hand, siamese structure makes DisIBS free from hand-craft features and combines this priori knowledge in data-driven way to find optimized features automatically. Both DisIBS and siamese structure have positive influence in small dataset learning. Finally, MLP classifiers map the discriminative features to accurate classification. Due to the discriminative features, it’s easy for MLP classifiers to classify HCC vs. ICC vs. Normal accurately. The Fig. 7 illustrates that the features extracted by SCCNN are discriminative for this task.

### E. EFFICIENCY

We compare SCCNN with other classic CNNs (vgg19_bn, inception_v3, resnet152) and CCNN in efficiency analysis. Specially, we use the convergence speed, training time and prediction time to measure the efficiency of algorithms. The train process is visualized to display convergence speed of different networks in Fig. 5.

As is shown in this figure, CCNN holds the fastest convergence speed and goes through about 22 epochs. resnet152 together with SCCNN shares the similar convergence speed, 42 epochs, and little slower than inception_v3. vgg19_bn is slowest to reach the convergence situation. We list the training time and prediction time information in Table 4.

| Network               | Training time | Params | Prediction time |
|-----------------------|---------------|--------|-----------------|
| vgg19_bn              | 112.378s      | 169.75M| 0.0086s         |
| resnet152             | 1114.12s      | 161.23M| 0.0414s         |
| inception_v3          | 735.62s       | 127.16M| 0.0390s         |
| CCNN                  | 490.06.86s    | 143.81M| 1.271s          |
| SCCNN(resnet152-based)| 2276.66s      | 161.23M| 0.0414s         |

As is shown in this figure, CCNN holds the fastest convergence speed and goes through about 22 epochs. resnet152 together with SCCNN shares the similar convergence speed, 42 epochs, and little slower than inception_v3. vgg19_bn is slowest to reach the convergence situation. We list the training time and prediction time information in Table 4.

| Network               | Training time | Params | Prediction time |
|-----------------------|---------------|--------|-----------------|
| vgg19_bn              | 112.378s      | 169.75M| 0.0086s         |
| resnet152             | 1114.12s      | 161.23M| 0.0414s         |
| inception_v3          | 735.62s       | 127.16M| 0.0390s         |
| CCNN                  | 490.06.86s    | 143.81M| 1.271s          |
| SCCNN(resnet152-based)| 2276.66s      | 161.23M| 0.0414s         |
FIGURE 6. Three filters’ outputs of the convolutional layer of two HCC images and two ICC image. The output feature maps with the size of $7 \times 7$ have been normalized to 0-255 in column wise. Next, a feature map is converted to the probability distribution vector by Equation (2), which is shown in Fig.7 (d), (e), (f) as symbol red, blue and green.

FIGURE 7. The probability distribution maps based on probability distribution vectors (Equation (2)) of 4 chosen images. Among them, (a) and (d) are the probability distribution maps of HCC and HCC before and after training. (b), (e) and (c), (f) are the probability distribution maps of ICC vs ICC, HCC vs ICC image respectively. In all figures, the red box represents filter 267, the green inverted triangle represents filter 1023 and the blue circle represents filter 2018. These discriminative filters are randomly selected.
same prediction time as resnet152 since they share the same structure in test process as mentioned above. CCNN uses a complex inference method and consumes much time.

F. VISUALIZATION

SCCNN extracts discriminative features based on the feature distribution. We visualize the feature distribution before and after training to verify the effectiveness of SCCNN. In this paper, we use a probability distribution map to describe the relation of two feature distributions. The horizontal axis and vertical axis of probability distribution maps are respectively the two probability distribution vectors (Equation (2)) of different images. As it is shown in Fig. 7 (a), (b), (c), almost all filters have no distinguish ability to classify liver cancers before training since the feature distribution are highly similar no matter whether the two images come from the same category or not. After training, feature distribution of two images from different categories (Fig. 7 (f)) obviously move away from the diagonal line, indicating that the filters learned some distinct structural features. We select 3 filters to show the convolutional results in Fig. 6. Taking filter 267 as an example, after training, 267th filter’s feature map of ICC has more pixels activated than the HCC image, which can also be observed in probability distribution scatter (Fig. 7 (f)). It means that filter 267 has learned a pattern which appears
more frequently in ICC. On the contrary, filters such as filter 2018 learned some patterns which appear more frequently in HCC, causing the points in distribution scatter move towards the left-up corner.

To try understanding the underlying mechanism of SCCNN, we followed Springenberg et al. [37] guided-backpropagation and Selvaraju et al. [38] grad-cam to visualize the activation of entire structure. Fig. 8 shows the results including the activation heatmap, Grad-CAM and guided-backpropagation results. We select randomly two HCC images, two ICC images together with two normal images in contrast. Grad-CAM shows the saliency distribution of SCCNN and highlights the regions of interest while guided-backpropagation reveals all the extracted features intuitively by gradients. A pathologist who has 10 years experience in hepatobiliary surgery annotates the regions of interesting on all the HCC and ICC images. What noteworthy is that the annotations are made copied with the raw dynamic CT files. Further analysis on the visualizations will be discussed in next section.

V. DISCUSSION
In this paper, we propose a novel framework named SCCNN which extends CCNN with the siamese mechanism and gradient-stop MLP classifiers. A new IBS-based measurement named DisIBS is designed for SCCNN to extract discriminative features. From the structural capability perspective, SCCNN is composed of two main parts including siamese feature extractors with DisIBS operator and MLP classifiers. Owing to DisIBS, SCCNN can fit the insufficient medical images and be insensitive to tumor size. Simase mechanism and gradient-stop MLP classifiers enable SCCNN to accept the random inputs and predict independently. SCCNN has four main characteristics:

- SCCNN uses siamese convolutional neural networks with a DisIBS-based metric function at the top as feature extractors and can accept random inputs for each pipeline.
- SCCNN is assigned different losses with different parts in train process. A DisIBS-based loss optimizes feature extractors to extract discriminative features while softmax-based loss only pushes MLP classifiers to derive the accurate classification.
- SCCNN adds gradient-stop MLP classifiers following feature extractors. Each one MLP classifier connected with the corresponding feature extractor can accept one example and output the classification result like traditional CNN in test process.
- SCCNN is easy to be adjusted by replacing the base model with other effective architectures.

We use SCCNN on a real clinical dataset composed of 234 persons and present the accuracy in slice/patient level. In this study, the highest slice/patient accuracy which we have achieved on three categories classification of HCC vs. ICC vs. Normal is 90.22%/94.92% and the accuracy rises to 94.17%/97.44% on two liver cancers classification, HCC vs. ICC. SCCNN achieves better performance than the original CCNN and other CNN methods in this task. SCCNN increases the binary and three-categories accuracy respectively by 11.6%/10.17%(slice/patient) and 8.39%/8.24%. Furthermore, there are three points to be further discussed.

Firstly, as is shown in Table 2, we find that the SCCNN's base model with residual blocks usually achieves better results such as resnet152, senet_resnet152 and densenet201. Residual blocks add the shallow information to the deeper layer and can preserve the bottom features as much as possible. We can conclude a possible thought that bottom features other than top features also play an important part in improving SCCNN. When compared with the resnet152-based one, the performance of senet_resnet152 version decreases much. This is possibly because the complexity of senet_resnet152-based one may be too large for this task. Another possible reason is that the attention mechanism may contradict with some assumptions of our framework. The specific cause of this phenomenon needs to be researched further.

Secondly, as can be seen in heatmap and Grad-CAM columns of Fig.8, the regions of interest for HCC and ICC images overlap with or include the annotations except the first HCC. Closer inspection of the first ICC results illustrates that SCCNN pays attention to not only annotated regions but also some other textures. When analyzing the guided-backpropagation views of HCC and ICC images, it’s apparent that the extracted features of SCCNN are diffuse and distributed in the whole liver instead of constrained in the lesion regions. For example, the guided-backpropagation visualizations of the second ICC activate almost the entire entity. All the results show that the details of images which SCCNN pays attention to exist not only in the lesion areas but also in other areas of the whole liver. When solving slices where tumor size is small or even unseen, the features in non-lesion areas may also help SCCNN infer the accurate results.

Thirdly, we segment the liver entity from the raw slice images to avoid the interference of extrahepatic tissue since the main task of this study is liver cancer classification. We use a simple semi-automated snake algorithm because it’s easy to implement and needs relatively few templates. However, it’s a coarse-gained segmentation and may influence the final classification accuracy. Some state of art medical segmentation methods will be tried in the following study.

VI. CONCLUSION
SCCNN extends CCNN with siamese mechanism and gradient-stop MLP classifiers, which makes it possible to accept the random inputs and predict like traditional CNN. With the new designed DisIBS, SCCNN can not only fit the insufficient medical images and be insensitive to tumor size but also extract more discriminative features. SCCNN shows a good performance in liver cancer classification on dynamic CT images. In practice, this method may be used as a computer-aided and noninvasive diagnosis that can help
doctors screen the patients from HCC vs. ICC vs. Normal by CT slices, which is hard to do directly by doctors’ eyes. Generally, SCCNN may offer a feasible choice for few data situations. This will be researched in the following study.

REFERENCES

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, “Global cancer statistics, 2012,” Cancer J. Clinicians, vol. 65, no. 2, pp. 87–108, Mar. 2015.

[2] W. Chen, R. Zheng, P. D. Baade, S. Zhang, H. Zeng, F. Bray, A. Jemal, X. Q. Yu, and J. He, “Cancer statistics in China, 2015,” Cancer J. Clinicians, vol. 66, no. 2, pp. 115–132, Mar. 2016.

[3] D. L. White, A. P. Thrift, F. Kanwal, J. Davila, and H. B. El-Serag, “Incidence of hepatocellular carcinoma in all 50 United States, from 2000 through 2012,” Gastroenterology, vol. 152, no. 4, pp. 812–820, Mar. 2017.

[4] B. Niel, Y. Rotman, I. Ditah, and J. K. Lim, “Emerging trends in hepatic carcinoma incidence and mortality,” Hepatology, vol. 61, no. 1, pp. 191–199, Jan. 2015.

[5] N. N. Massarweh and H. B. El-Serag, “Epidemiology of hepatocellular carcinoma and intrahepatic cholangiocarcinoma,” Cancer Control, vol. 24, no. 3, Sep. 2017, Art. no. 1073274817792294.

[6] Y. Wei, F. Gao, D. Zheng, Z. Huang, F. Bray, A. Jemal, X. Q. Yu, and J. He, “Cancer statistics in China, 2015,” Cancer J. Clinicians, vol. 65, no. 2, pp. 115–132, Mar. 2016.

[7] S. Tsunematsu, M. Chuma, T. Kamiyama, N. Miyamoto, S. Yabusaki, A. S. Ashour, Y. Guo, A. R. Hawas, and G. Xu, “Ensemble of subspace discriminant classifiers for schistosomal liver fibrosis staging in mice microscopic images,” Health Inf. Sci. Syst., vol. 6, no. 1, p. 21, Dec. 2018.

[8] Y. Wang, Y. Chen, N. Yang, L. Zheng, N. Dey, A. S. Ashour, V. Rajinikanth, J. M. R. S. Tavares, and F. Shi, “Classification of mice hepatic granuloma microscopic images based on a deep convolutional neural network,” Appl. Soft Comput., vol. 74, pp. 40–50, Jan. 2019.

[9] A. Midya, J. Chakraborty, L. M. Pak, J. Zheng, W. R. Jarnagin, R. K. Do, and A. L. Simpson, “Deep convolutional neural network for the classification of hepatocellular carcinoma and intrahepatic cholangiocarcinoma,” Proc. SPIE, vol. 10575, Feb. 2018, Art. no. 1057528.

[10] Y. Huang, Y. Chen, H. Zhu, W. Li, Y. Ge, X. Huang, and J. He, “A liver fibrosis staging method using cross-contrast network,” Expert Syst. Appl., vol. 130, pp. 124–131, Sep. 2019.

[11] A. C.-C. Yang, S.-S. Hsu, H.-W. Yen, A. L. Goldberger, and C.-K. Peng, “Linguistic analysis of the human heartbeat using frequency and rank order statistics,” Phys. Rev. Lett., vol. 90, no. 10, Mar. 2003, Art. no. 108103.

[12] J. Bromley, J. W. Bentz, L. Bottou, I. Guyon, Y. LeCun, C. Moore, E. Säckinger, and R. Shah, “Signature verification using A Siamese time delay neural network,” IJIPRAI, vol. 7, no. 4, pp. 669–688, 1993.

[13] M. Amin-Najii, H. Mahdavianiat, and A. Aghapozadheh, “Alzheimer’s disease diagnosis from structural MRI using siamese convolutional neural network,” in Proc. 4th Int. Conf. Pattern Recognit. Image Anal. (IPRIA), Mar. 2019, pp. 75–79.

[14] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Deep metric learning for person re-identification,” in Proc. 22nd Int. Conf. Pattern Recognit., Stockholm, Sweden, Aug. 2014, pp. 34–39.

[15] K. Chen and A. Salman, “Extracting speaker-specific information with a regularized Siamese deep network,” in Proc. 24th Int. Conf. Neural Inf. Process. Syst. (NIPS), Red Hook, NY, USA: Curran Associates, 2011, pp. 298–306.

[16] Y. Sun, Y. Chen, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” in Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS), vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 1988–1996.

[17] A. L. Goldberger and C.-K. Peng, “Genomic classification using an information-based similarity index: Application to the SARS coronavirus,” J. Comput. Biol., vol. 12, no. 8, pp. 1103–1116, Oct. 2005.

[18] C.-K. Peng, A. C.-C. Yang, and A. L. Goldberger, “Statistical physics approach to categorize biologic signals: From heart rate dynamics to DNA sequences,” Chaos, Interdiscip. J. Nonlinear Sci., vol. 17, no. 1, Mar. 2007, Art. no. 015115.

[19] L. D. Cohen, “On active contour models and balloons,” CVGIP: Image Understand., vol. 53, no. 2, pp. 211–218, Mar. 1991.

[20] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, arXiv:1409.1556. [Online]. Available: http://arxiv.org/abs/1409.1556

[21] S. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2818–2826.

[22] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[23] S. Xie, R. Girshick, P. Dollar, and K. He, “Aggregated residual transformations for deep neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1492–1500.

[24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4700–4708.

[25] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.

[26] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for simplicity: The all convolutional net,” 2014, arXiv:1412.6806. [Online]. Available: https://arxiv.org/abs/1412.6806

[27] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, Oct. 2017, pp. 618–626.
QIYUAN WANG received the B.S. degree from Nanjing University, China, in 2018, where he is currently pursuing the master’s degree. His research interests include machine learning and medical images analysis.

ZHONGMIN WANG received the M.S. and Ph.D. degrees. He is currently a Nanjing University and Southeast University part-time master’s Tutor. He has completed a number of provincial and municipal projects at the national level, published dozens of articles in core journals at home and abroad, and won two national invention patents and two national academy awards.

YU SUN received the B.S. degree from the Nanjing University of Posts and Telecommunications, China, in 2016, where he is currently pursuing the master’s degree in biomedical engineering. His research interest is application of machine learning in medical images.

XIN ZHANG received the M.S. degree in biomedical engineering from Southeast University, China, in 2008. She obtained the title of Senior Engineer, in 2016. She is mainly engaged in medical informatics research and hospital informatization construction, especially in medical big data analysis and medical image processing.

WEIFENG LI received the Ph.D. degree in biomedical engineering from Southeast University, China, in 2018. He is currently a Researcher with Nanjing University. His research interests include medical devices, medical physics, medical image, and image processing.

YUN GE received the Ph.D. degree in biomedical engineering from Southeast University, China, in 2001. He is currently a Professor with Nanjing University. His research interests include medical devices, medical physics, medical image, and image processing.

XIAOLIN HUANG received the Ph.D. degree in acoustics from Nanjing University, China, in 2009. She is currently an Associate Professor with Nanjing University. Her research interests include detection, processing, and analysis of biomedical signals.

YUN LIU received the Ph.D. degree. She is currently a Professor and the Chief Physician. She is also the Vice President of the Jiangsu Provincial People’s Hospital, a Doctoral Supervisor of Nanjing Medical University, the Director of the Institute of Medical Informatics and Management, Nanjing Medical University, and the Vice President of the School of Biomedical Engineering and Information, Nanjing medical university.

YING CHEN received the Ph.D. degree in acoustics from Nanjing University, China, in 2003. He is currently an Associate Professor with Nanjing University. His research interests include biomedical signal processing and image processing. His current research interest includes disease detection using deep learning method.