Communications in Information and Systems
Volume 20, Number 4, 415–442, 2020

CasCRNN-GL-Net: cascaded convolutional and recurrent neural networks with global and local pathways for classification of focal liver lesions in multi-phase CT images

Dong Liang*, Yingying Xu*, Lanfen Lin†, Nan Zhou, Hongjie Hu†, Qiaowei Zhang, Qingqing Chen, Xianhua Han, Yutaro Iwamoto, and Yen-Wei Chen†

Automatic focal liver lesion (FLL) classification based on multi-phase computed tomography (CT) images is one of the most crucial components of a computer-aided diagnosis (CAD) system for liver disease. Though much research has been conducted in this field, two challenges still remain: (1) difficulty of representing the temporal enhancement pattern effectively and (2) the need to obtain the local and global information of FLLs. However, most of existing studies only focus on local or global information. In this paper, we propose a cascaded convolutional neural network (CNN) and recurrent neural network (RNN) with global and local pathways, called CasCRNN-GL-Net, for the classification of FLLs in multi-phase CT images. The CNN with global and local pathways is used to extract both global and local information. A bi-directional long short-term memory (BD-LSTM) model is employed to represent the temporal enhancement pattern in multi-phase CT images. Moreover, we propose a novel joint loss function to train the proposed framework. The joint loss function is composed of inter-class and intra-class losses, which can improve the robustness of the framework. The proposed method outperformed state-of-the-art approaches by achieving a mean accuracy of 87.64%. We have released the code in the GitHub (https://github.com/UpCoder/GL_BD_LSTM).

1. Introduction

Liver cancer is the second and sixth most common cause of cancer-related deaths among men-, and women worldwide, respectively. Moreover, hep-

*These authors contributed equally to this work.
†Co-Corresponding Authors: Lanfen Lin, Hongjie Hu and Yen-Wei Chen.

atocellular carcinoma (HCC) is the most common type of primary liver cancer [1]. In addition to HCC, other types of focal liver lesions (FLLs) include focal nodular hyperplasia (FNH), Cyst, hemangiomas (HEM) and metastasis (METS).

Radiological examinations are the primary methods for diagnosing FLLs in clinical. However, tumor imagery assessments are generally subjective and rely on the experience of radiologists. Computer-aided diagnosis (CAD) systems which eliminate the influence of subjectivity are able to assist radiologists in clinical diagnosis. Automatic classification of FLLs is one of the most crucial components of a CAD system for diagnosis of liver tumors.

Currently, multi-phase contrast-enhanced computed tomography (CT) images, which are also known as dynamic CT images, are the most crucial imaging modality employed to detect, locate and diagnose FLLs. Multi-phase CT scans are generally divided into four phases, namely, non-contrast (NC) phase (before contrast injection), arterial (ART) phase (30–40 s after contrast injection), portal venous (PV) phase (70–80 s after contrast injection), and delay (DL) phase (3–5min after contrast injection). The appearance and temporal changes of FLLs in multi-phase CT images are important in diagnosing the type of FLL. Fig. 1 shows the typical images of five FLLs over the three phases (NC, ART and PV. DL phase is not provided as DL is optional in clinical diagnosis). The characteristics of different types of FLLs are listed as follows: (1) Cysts do not show any enhancement after intravenous contrast injection. (2) HEM exhibits discontinuous nodular peripheral enhancement in the ART phase with centripetal enhancement over time. Calcification and central fibrosis may sometimes be observed because of thrombosis in the vascular channels, thus making it challenging to discriminate HEM. (3) FNH is usually hypo- or isodense to the liver parenchyma without any contrast. FNH demonstrates bright ART contrast enhancement except for the central scar and pronounced central arteries may be visible. In the PV phase, FNH becomes isodense to the liver tissue. HEMs and FNHs show similar peripheral enhancement in the ART phase. The presence of central abnormalities sometimes makes the PV washout of HEM and FNH visually similar. (4) HCC typically shows hyperenhancement in ART phase and washout in PV phase. (5) On the contrary, METS is a malignant tumor that usually spreads from other cancer affected organs. Although METS are enhanced homogeneously, they tend to have less well-defined margins than cysts [2]. Moreover, we have observed that when human experts diagnose the type of FLLs, they tend to zoom in of the CT images to analyze the detail of the lesions [3]. The experts also need to look back or forward in different phases, and this approach can help doctors recognizing the type of

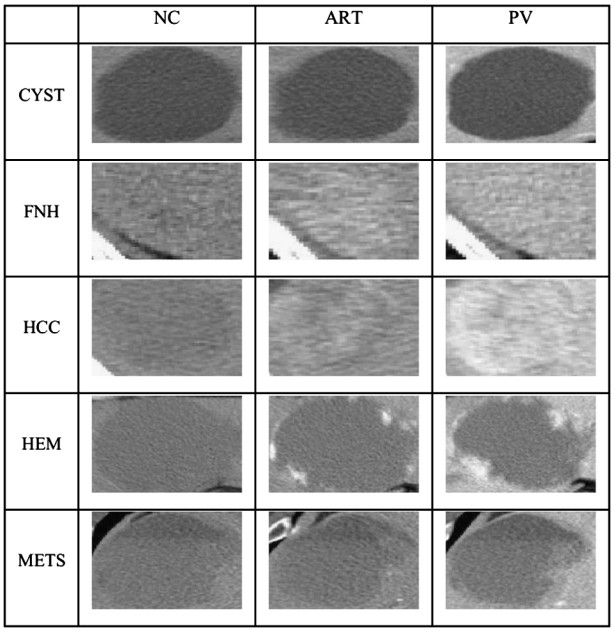


Figure 1: Typical images of five liver lesions over the three phases.

tumor. This observation indicates the importance of combining local with global information with the temporal enhancement pattern.

1.1. Related work

Some studies have reported the characterization of FLLs using multi-phase images to capture the spatial and temporal information among phases. We summarized them as three groups:

A1. Low-level features: Roy et al. [2] proposed a framework extracting spatiotemporal features from multi-phase CT images to characterize FLLs. The spatiotemporal features inculude density features (the normalized average intensity of a lesion), texture features (the gray-level co-occurrence matrix [GLCM] which includes nine types of texture coefficients), temporal density and texture features (the intensity and texture over three enhancement phases compared with the non-contrast phase). Note that this is a typical method that only distills the low-level features.

A2. Mid-level features: Compared with low-level features, mid-level features such as bag-of-visual-words (BoVW) [4] and its variants have proved to be considerably effective in classifying FLLs [5, 6, 7, 8, 9]. The BoVW

is an extensively used feature-modeling approach that can integrate various low-level features like intensity, texture, and spatial information. It clusters a low-level feature space into a number of regions that potentially correspond to visual concepts which are called visual words.

Yang et al. [5] employed the BoVW model to construct a visual category-specific vocabulary for each pathological class.The histograms of each single phase were merged together to represent multiphase images. Xu et al. [6] proposed a texture-specific BoVW method to represent FLLs. Pixels in the regions of interest (ROIs) are classified into nine texture categories using the rotation-invariant uniform local binary pattern method. Thereafter, the BoVW-based features were calculated for each texture category, the concatenation operation was performed. Furthermore, a spatial cone matching (SCM) based representation strategy is proposed to describe the spatial information of the visual words in the ROI.

Yu et al. [7] improved the original BoVW method, which builds the dictionary by using multi-ROIs to extract the extra information. They asserted that the density discrepancy between the adjacent normal liver parenchyma and tumors in triple-phase CT images is also a foundation for the diagnosis of radiologists. Therefore, they split the surrounding liver parenchyma of lesions into four different regions. Diamant et al. [8] also proposed a dual-dictionary to represent the boundary and interior regions, respectively. After dictionary generation, the two histograms are concatenated to represent the tumors.

Xu et al. [9] proposed an algorithm based on BoVWs to represent the feature of liver tumors and their retrieval. To distill the enhancement pattern, the authors used registered triple-phase images as a color-like input images and used the color-like images (the triple-phase images) to constructed a temporal co-occurrence vocabulary. Wang et al. proposed a sparse coding based BoVW [10] and a tensor sparse coding based BoVW [11] to realize soft assignment of codebooks for effective representations of FLLs in multi-phase CT images.

A3. High-level features: The high-level feature representation of deep convolutional neural networks (DCNNs) has achieved great successes in various challenging tasks, including in the field of medical image analysis. However, owing to the limitation of the scale of the dataset, only a few studies with deep learning method on the classification of focal liver lesions.

Frid-Arar et al. [12] tackled the FLL detection task as a patch-based classification task based on a multi-scale patch-based classification framework. They only used the single-phase (PV phase) as the input of networks. Moreover, different from the conventional two classes, they proposed triple-classes

(lesion, normal-interior and normal-boundary) to improve the performance. They also proposed two parallel stream architectures to combine different types of scale information.

Yasaka et al. [13] proposed a convolutional neural network with three channels corresponding to three phases (NC, ART, and DL) to classify liver tumors in dynamic contrast-enhanced CT images. The method involves six convolutional layers, three maximum pooling layers and three fully connected layers. The features extracted from each single phase were concatenated to classify FLLs.

The limitations of the above-mentioned methods are as follows: (1) The feature concatenation of each single phase lacks information on inter-phase enhancement pattern (temporal enhancement information and relationship among phases). (2) They focus on either global features or local features. However, as mentioned previously, the coarse and fine features are both important for FLL diagnosis.

1.2. Contributions

In this paper, we proposed a framework based on cascaded convolutional neural network with global and local pathways (CNN-GL-Net) and recurrent neural network (bi-directional long short-term memory (BD-LSTM)) model), which is called as CasCRNN-GL-Net, for the classification of FLLs. A preliminary version of this work has been presented at a conference [14]. The main contributions of this paper are summarized as follows:

- We extracted features from each single-phase CT image by using CNN-GL-Net. The input of CNN-GL-Net is a pair (patch and ROI) that represents the local and global information to handle inter-class similarities.
- We captured the enhancement pattern that is hidden in multi-phase CT images by using BD-LSTM block to represent each patch. To the best of our knowledge, expressing temporal features (enhancement patterns) among multiphase images using deep learning has not been investigated previously.
- We proposed a new joint loss function to train our model, and provide a more robust and accurate deep model. The loss function is composed of an inter-loss and intra-loss, which minimizes the inter-class and intra-class variations, respectively and we updated the center value by using a back-propagation process.
- Extensive experiments with promising results have been validated on multi-phase abdominal CT images.

- Compared to our conference paper [14], which is the basis of this paper, the main extended parts are shown in below: (1) We add the extra geometry information of lesion (e.g. size, roundness, etc.) to improve the result. (2) The types of FLLs is increased from four to five. In addition to existing Cyst, HCC, HEM and FNH, we add a new type of FLL (i.e., METS) in the experiments. (3) We use a pretrained model to improve the classification accuracy. (4) Detailed description and discussion about the proposed method and experimental results are added in this paper.

2. Methodology

Fig. 2 shows an overview of the proposed framework. In Section 2.1, we first show how ROIs and patches are prepared, which are used as input pairs of the proposed network. In Section 2.2, we describe the proposed deep learning architecture, called CasCRNN-GL-Net, which combines a deep convolutional neural network with global and local pathways (CNN-GL-Net) with a bi-directional long short-term memory (BD-LSTM) model for the pixel-wise classification of FLLs in multi-phase CT images. The proposed joint loss function, which is composed of an inter-loss and intra-loss, is presented in Section 2.3. In Section 2.4, we describe the features extracted from the label map (pixel-wise classification result), and how we accomplish the lesion-based classification. Sections 2.2 and 2.3 are core techniques of the proposed framework. Sections 2.1 and 2.4 can be considered as a pre-processing and a post-processing step, respectively.

2.1. Pre-processing (data preparation)

The input of the proposed network is a pair of patch and ROI. First, we applied a random-walk based interactive segmentation algorithm [15] to segment the liver (healthy tissues) and FLLs. The segmented results were checked by two experienced radiologists. The segmentation was performed for each phase image separately. One example of CT image (PV phase) and its segmented lesion are shown in Figs. 4(a) and 4(b). Each ROI has a label corresponding to its diagnosis result, i.e. $c \in \{c_0, c_1, c_2, c_3, c_4\}$ where c_0 represents cyst, c_1 represents FNH, c_2 represents HCC, c_3 represents HEM, and c_4 represents METS. During a clinical CT study, the spatial placement of tissues formed in multiple phases exhibits some aberration owing to differences in the patient's body position, respiratory movements, and heartbeat.

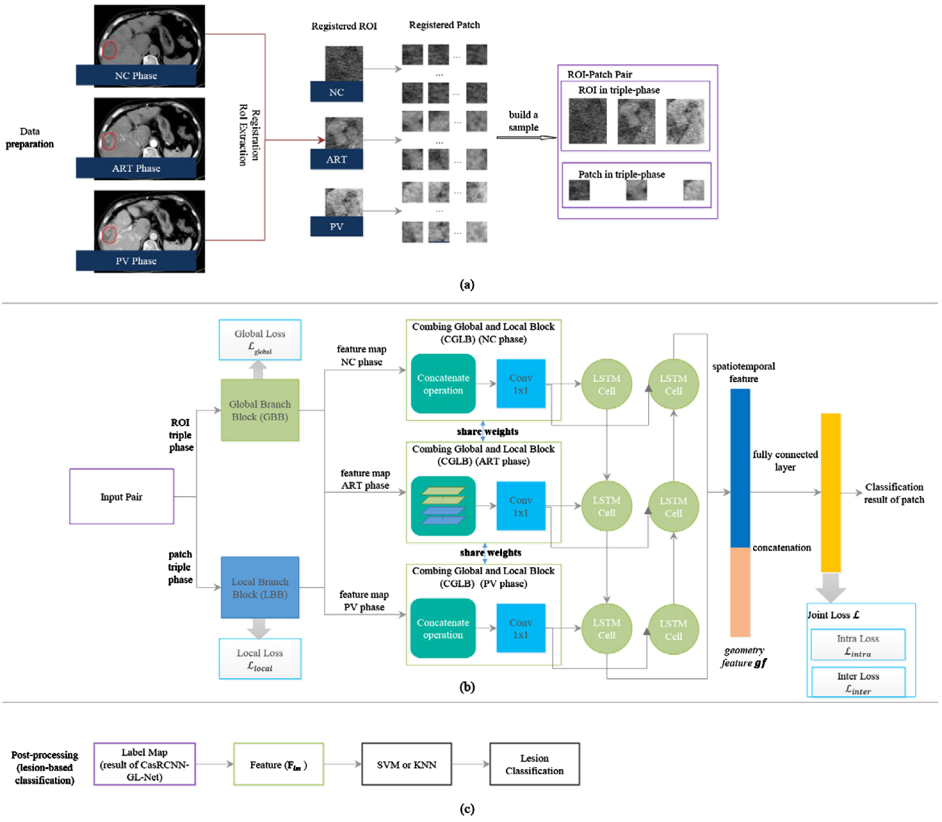


Figure 2: Overview of the proposed framework for classification of FLLs based on multi-phase CT images. (a) shows the data preparation process. We first performed the registration operation among the triple-phase images. Thereafter, we extracted the patches by using a patch-extracted method. Finally, we used the pair of patch and ROI as an input (a sample). (b) shows the architecture of the proposed network (CasCRNN-GL-Net). The global branch block (GBB) and local branch block (LBB) are followed by BD-LSTM operation to distill the global spatiotemporal feature (GSF) and local spatiotemporal feature (LSF), respectively. The GSF and LSF are used to compute the global branch loss and local branch loss. (c) shows the post-processing for lesion-based classification based on the normalized label map (result of CasCRNN-GL-Net).

Therefore, to obtain a factual variation of the density over phases, we applied a non-rigid registration technique to localize a reference lesion in other phases [16]. Each segmented lesion image (i.e., 2D slice image) was resized to 128×128 pixels. The resized images were then used as the input for global pathway training and testing.

Patches were then extracted from ROIs. Each patch has the same label with the ROI. Owing to significant variations in the sizes of the different lesions, extreme imbalances occur among the patch categories. To solve this problem, we propose to use a pace value for patch extraction. The pace value is derived in Eq. (1):

$$pace_i = \begin{cases} floor(\sqrt{\frac{(w_i * h_i)}{\epsilon}}), & w_i * h_i \geq \epsilon \\ 1, & else \end{cases} \tag{1}$$

where i represents the i-th ROI; $pace_i$ represents the pace of i-th ROI for extracting the patches; w_i and h_i represent the width and height of the i-th ROI, respectively. ϵ represents a threshold that can limit the number of patches. The floor function represents a rounding-down operation. ϵ was set to 128 as a common practice. For the testing dataset, we still set the pace to one. Similar to the global pathway approach, we resized the patches to 64×64.

2.2. Deep learning architecture: CasCRNN-GL-Net

Section 2.2.1 describes the CNN-GL-Net in detail which involves a global branch block (GBB) and a local branch block (LBB), that extract global and local information (as well as the intra-phase features), respectively, from every single phase. The global and local information are fused using the GL fusion block. Section 2.2.2 describes the BD-LSTM block, which captures the enhancement pattern over phases as well as the inter-phase features. Section 2.2.3 describes the method combining the CNN-GL-Net block and BD-LSTM block. We will also introduce the method to combine high-level features (i.e. BD-LSTM output) with the geometry features in Section 2.2.4.

2.2.1. CNN-GL-Net In this subsection, we describe the CNN-GL-Net block, which was proposed in our previous work [17]. The CNN-GL-Net block comprises the global branch block (GBB), local branch block (LBB), and the combining global and local block (CGLB). The GBB involves three base CNNs which are used to extract features from NC, ART and PV phases, respectively. The three base CNNs with the same architecture, which distill

the basic visual feature, share weights with each other [18, 19]. It should be noted that we can use any deep CNNs to extract the feature from the ROI. For example, we can use the ResNet-50, which is shown in Fig. 3, to implement the deep CNNs. The LBB architecture is similar to the GBB, thus, we will not repeat its discussion.

The CGLB is shown in Fig. 2(b). First, we conducted a concatenate operation to combine the local feature maps and global feature maps. We then distilled the intra-phase feature map by using a convolutional layer with a 1×1 kernel.

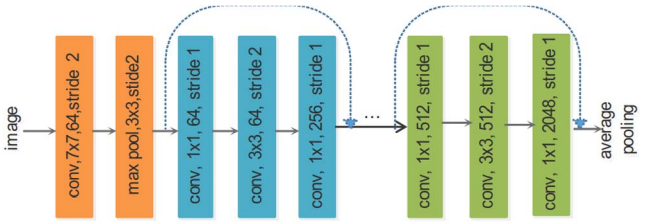


Figure 3: The architecture of ResNet-50.

2.2.2. BD-LSTM A recurrent neural network (RNN) can maintain a self-connected status and act as a memory to remember previous information as sequential data. Long-short term memory (LSTM) [20] is a class of RNN that can avoid the vanishing gradient problem.

Bi-directional LSTM (BD-LSTM), which stacks two layers of LSTM, is an extension of LSTM. The two layers of LSTM, which are illustrated in Fig. 2(b), work in two opposite directions to extract temporal information from sequential data. The enhancement information carried in the two layers of LSTM is concatenated as the output. One layer is in the t^+-direction and extracts the enhancement pattern from the NC phase to the PV phase. The other layer is in the t^--direction and extracts the anti-enhancement pattern from the PV phase back to the NC phase.

2.2.3. Combining CNN-GL-Net and BD-LSTM The motivation of performing FLL classification based on multi-phase CT images by combining CNN-GL-Net and BD-LSTM is to employ multi-phase CT images as sequential data. The CNN-GL-Net extracts the information (i.e., intra-phase feature) on the basis of a single phase. The BD-LSTM distills the enhancement information (i.e., inter-phase feature) among three phases. The length

of sequential data is a constant number (i.e., three). The two blocks work in coordination, as follows.

The output of the three CNN-GL-Net blocks, as a sequential data, constitutes the input of the BD-LSTM. Furthermore, the output of the two layers of LSTM (i.e., BD-LSTM) constitutes the input of the fully connected layer. Moreover, its output represents the high-level features of patches including the spatiotemporal features.

2.2.4. Combining the high-level features and the geometry information We have introduced in the previous subsection the method to extract high-level features to represent the patch. However, we resized the patch and ROI (as well as lesion) to 64×64 and 128×128, respectively, by ignoring the size of the tumor or other geometry information that is considered important. Hence, to utilize the geometry information, we extracted the width, height, size, perimeter, and roundness of each phase as complementary information. Formally, **gf** (geometry feature vector) is defined as follows:

$$\mathbf{gf} = \{\mathrm{gf_{nc}}, \mathrm{gf_{art}}, \mathrm{gf_{pv}}\} \tag{2}$$

$$\mathrm{gf_{nc}} = (w_{nc}, h_{nc}, s_{nc}, p_{nc}, r_{nc}) \tag{3}$$

$$w_{nc} = x_{max} - x_{min} \tag{4}$$

$$h_{nc} = y_{max} - y_{min} \tag{5}$$

$$r_{nc} = \frac{4 * \pi * s_{nc}}{p_{nc}^2} \tag{6}$$

where $\mathrm{gf_{nc}}$, $\mathrm{gf_{art}}$ and $\mathrm{gf_{pv}}$ represent the geometry feature vector of the NC, ART and PV phase CT images, respectively. Eq. (3) shows the definition of $\mathrm{gf_{nc}}$. The definition of $\mathrm{gf_{art}}$ and $\mathrm{gf_{pv}}$ is similar to $\mathrm{gf_{nc}}$; thus we will not repeat their definition. Eqs. (4) and (5) show the definition of the width and height, respectively. The x_{max}, x_{min}, y_{max} and y_{min} are the upper left and lower right vertex coordinates of the bounding box (ROI). s_{nc} and p_{nc} represent the number of pixels in the tumor region and tumor boundary, respectively. r_{nc} represents the roundness, and is defined in Eq. (6).

We then conducted the concatenate operation to combine the high-level features and geometry features. The fully connected layer was used to extract the feature to represent the patch automatically. The softmax layer following the last fully connected layer produces an output that provides the result of the patch-based classification (assigning a class label to the center pixel of the patch).

2.3. Joint loss function

In this section, we will introduce the proposed joint loss function and training strategy of the framework.

Let N be the batch size and ω^t be the weight in the t-th ($t = 1, 2, \ldots, T$) layer. We used $\boldsymbol{W}$ to denote the weight of the mainstream network (Fig. 2), where $\boldsymbol{W_{local}}$ and $\boldsymbol{W_{global}}$ represent the weights of the local and global pathways, respectively. Note that the GBB and LBB is followed by BD-LSTM operation to distill the global and local spatiotemporal features (GSF, LSF), respectively. The GSF and LSF are used to compute the $\mathcal{L}_{local}$ and $\mathcal{L}_{global}$, respectively. Furthermore, $p(j|x_i; \boldsymbol{W})$ represents the probability of the i-th patch belonging to the j-th class. We define $p(j|x_i; \boldsymbol{W_{local}})$ and $p(j|x_i; \boldsymbol{W_{global}})$ similarly. The definitions of cross-entropy are as follows:

$$\mathcal{L}_{mainstream} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} -p(j|x_i; \boldsymbol{W}) * \log(p(j|x_i; \boldsymbol{W})) \tag{7}$$

$$\mathcal{L}_{local} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} -p(j|x_i; \boldsymbol{W_{local}}) * \log(p(j|x_i; \boldsymbol{W_{local}})) \tag{8}$$

$$\mathcal{L}_{global} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{C} -p(j|x_i; \boldsymbol{W_{global}}) * \log(p(j|x_i; \boldsymbol{W_{global}})) \tag{9}$$

where the C represents the number of categories. The definition of the inter-loss is shown as follows:

$$\mathcal{L}_{inter} = \frac{1}{2} * \mathcal{L}_{mainstream} + \frac{1}{4} * \mathcal{L}_{local} + \frac{1}{4} * \mathcal{L}_{global} \tag{10}$$

In addition to the above inter-loss (Eq. (10)), we also define an intra-loss (i.e. center loss [21]) to treat the inter-variation problem, which is defined as follows:

$$\mathcal{L}_{intra} = \frac{1}{2} \sum_{i=1}^{N} \left|\left| \boldsymbol{f}_i - \boldsymbol{c}_{\boldsymbol{y_i}} \right|\right|^2 \tag{11}$$

where $\boldsymbol{f}_i$ is the representation feature of the i-th patch and $\boldsymbol{c}_{y_i}$ denotes the y_i-th class center of features. In the course of training $\boldsymbol{c}_{y_i}$ should be updated using the back-propagation process. To accelerate our training, we conducted the update operation on the base of each batch instead of basing

it on the entire training set. In this case, some of the centers may not change. The method employed to update the centers is described as follows:

$$\begin{aligned} \boldsymbol{c_j} &= \boldsymbol{c_j} - \boldsymbol{\Delta c_j} = \boldsymbol{c_j} - \frac{\boldsymbol{\partial \mathcal{L}}}{\boldsymbol{\partial c_j}} \\ &= \boldsymbol{c_j} - \alpha * \frac{\sum_{i=1}^{N} \delta(y_j == j)(\boldsymbol{f_i} - \boldsymbol{c_j})}{1 + \sum_{i=1}^{N} \delta(y_i == j)} \end{aligned} \tag{12}$$

where $\delta(y_i == j) = 1$ if $y_i == j$ holds, otherwise, $\delta(y_i == j) = 0$. Furthermore, α can restrict the learning rate of the centers, where the range of α is $(0, 1)$.

Finally, we adopted a joint loss that combines the intra-loss and inter-loss to train the framework. Eq. (13) shows the formulation of the joint loss. We used λ to tackle the balance between the inter-class loss and intra-class loss.

$$\mathcal{L} = \mathcal{L}_{inter} + \lambda \mathcal{L}_{intra} \tag{13}$$

Our framework is split into two phases comprising the training phase and testing phase. During training, we first trained the part of our framework that involves the deep learning components and then aggregated the label maps. The effectiveness of the patches from the validation dataset determines when the training should stop. We aggregated the label maps that belong to the training and validation datasets after the model was trained. We then used the training and validation label maps as input to the support vector machine (SVM) classifier for final lesion-based classification (Fig. 2(c)). We also determined the parameters of the SVM (classifier of lesions) by using the effectiveness of the label map of the validation dataset.

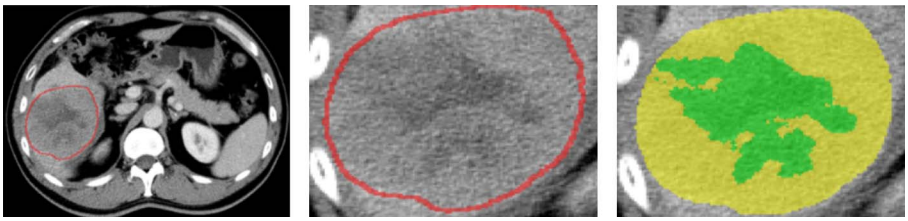
2.4. Post-processing (lesion-based classification)

After pixel-wise classifier training, we aggregated the label map of each lesion. For showing the label map more intuitively, one example is shown in Fig 4(c). In the label map, each pixel was assigned with a class label based on the pixel-wise classification (Fig. 2(b)), which is represented by different color. Figs. 4(a) and 4(b) are the CT image (PV phase) and its ROI (segmented lesion), respectively. Then, we extracted features Flm from the label map as follows:

$$\mathrm{F_{lm}} = \{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\} \tag{14}$$

where β_j, which is derived in Eq. (15), denotes the proportion of pixels belonging to the j-th category of the label map. We then used the SVM to achieve lesion-based classification.

$$\beta_j = \frac{\text{the number of pixels belong to } j\text{-th category}}{\text{the total pixels in the label map}} \tag{15}$$



(a) (b) (c)

Figure 4: (a) represents the CT image slice (PV phase). (b) represents the ROI (segmented lesion), which is checked by two experienced experts. (c) represents the label map, in which each pixel was assigned with a class label represented by different color.

3. Experiments

3.1. Dataset

A total of 573 CT liver slice images were collected from Sir Run Run Shaw Hospital, containing five types of lesions confirmed by pathologists, (i.e., Cyst, FNH,HCC, HEM and METS). The slices were obtained from 302 cases. We selected 1∼5 slices with the significant appearance from each case for our experiments. The CT images in our dataset are abdominal CT scans taken from 2015 through 2017. The CT scans were acquired with a slice collimation of 5–7 mm, a matrix of 512×512 pixels, and an in-plane resolution of 0.57–0.89 mm. In our experiment, we randomly split our 302 CT scans (cases) into a training dataset, a validation dataset, and a testing dataset. The ratio of 2D slice images in the training, validation and test dataset is about 3:1:1. In order to eliminate the effect of randomness, we conduct the partition operation twice, and form two groups of dataset. The distribution of our dataset is shown in Table 1. The number in the parentheses represents the number of cases.

Table 1: The distribution of 2D slice images used for our experiments. The distribution of cases is represented in parenthesis

Type	Cyst		FNH		HCC		HEM		METS	
	Set1	Set2	Set1	Set2	Set1	Set2	Set1	Set2	Set1	Set2
Training	61(22)	69(28)	71(27)	60(24)	75(46)	69(39)	62(40)	79(38)	49(16)	68(19)
Validation	23(18)	24(16)	25(18)	31(20)	31(18)	27(15)	36(19)	28(28)	16(6)	7(7)
Testing	26(20)	17(16)	18(20)	23(21)	26(12)	36(22)	26(14)	17(16)	28(10)	18(6)
Total	110(60)		132(76)		93(32)		124(73)		93(32)	

We adopted a normalized strategy for the geometry feature, which will be normalized to [0, 1] as the input.

For augmentation of the ROI and patch image, we adopted the rotate and flip strategies. The images will be rotated at a probability of 0.5, at random angles of 0, $\pi/2$, π and $3\pi/2$. The images will be flipped horizontally and vertically at a probability of 0.5. The scale of the data has been expanded eightfold to 4584.

3.2. Implementation details

Our framework was implemented using the Tensorflow library in Python 2.7. First, we loaded the pre-trained model parameters trained by ImageNet [22]. Thereafter, we initialized the uninitialized parameters by using the xavier initializer [23]. We used a momentum optimizer to update our parameters by setting the initial learning rate to 1e-4 and the momentum coefficient to 0.9. We applied a polynomial decay to the learning rate; the end learning rate is set to 1e-5 with twenty thought steps. We set the batch size to 100. We evaluated the performance of our model by using the mean and standard deviation of the two testing set accuracies. We published the code in the GitHub.[1]

3.3. Parameter optimization

The patch size, and λ have an effect on the classification result. Thus, we varied them in our experiments to achieve better performance.

3.3.1. Patch size First, we set the patch sizes to 3, 5, 7 and 9. The validation accuracy of the patch will determine which patch size will be used. Fig. 5 shows the mean accuracy on the two datasets. Moreover, we can ob-

[1] https://github.com/UpCoder/GL_BD_LSTM

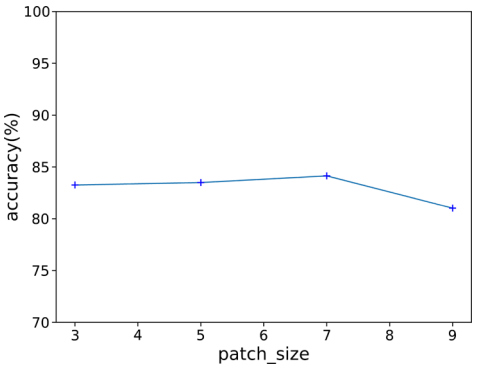


Figure 5: The accuracy with different patch size.

serve that as the size of a patch increases, the accuracy will also increase first and then decrease because the information of a patch is balanced between the local and global information. The trade-off will lead to the decrease in accuracy as the patch size increases. Note that a 7×7 patch size generates a better classification performance than others. Hence, we set the patch size to seven. When we performed the experiments, we fixed α to 0.5 and λ to 1.0.

3.3.2. Lambda and alpha The hyperparameter α controls the learning rate of the intra-loss in our model. The hyperparameter λ dominates the weight of intra-loss. Both parameters are essential to our method. Hence, we conducted two group experiments to investigate the sensitiveness of the two hyper parameters.

In the first experiment, we fixed λ to 1.0 and varied α from 0.0 to 1.0 to compare their effectiveness. The verification of the accuracy of the different α is illustrated in Fig. 6(a), which shows that when α does not update the center feature (α is zero), more noise will be incurred in the training process. We observed that the verification performance remains stable across a wide range of α.

In the second experiment, we fixed α to 0.25 and varied λ from 0.0 to 1.5 to explore their effectiveness in different λ. The verification of the accuracy of these different λ values is illustrated in Fig. 6(b), which shows that simply using the inter-loss ($\lambda = 0$) is not a good choice. We also observed that the performance of our method remains largely stable across a range of λ.

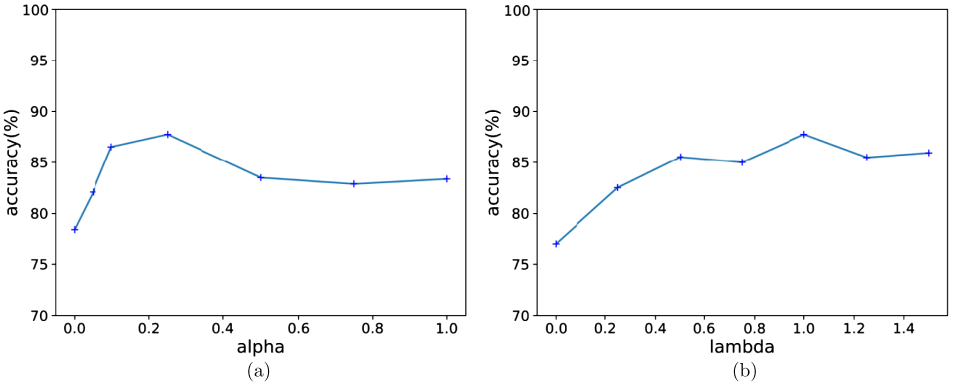


Figure 6: The accuracy with different α and λ.

3.4. Comparison with different base networks

In this subsection, we compared the effectiveness on the basis of different base networks. The ResNet [24] and VGG16 [25] have achieved great success in many fields. Hence, we used ResNet50 and VGG16 as base networks. We used KNN or SVM, which is simple but effective for our task, as the lesion-level classifier (Fig. 2(c)) to compare the mean accuracy and standard deviation. Table 2 shows the results. We observed that ResNet50 achieved better performance than VGG16, and ResNet50 was more stable (smaller standard deviation) than VGG16. Furthermore, SVM and KNN, which are two types of lesion-level classifiers, achieved similar performance. We deduced that the observation was due to the simplicity of the label map feature $\mathbf{F}_{\text{lm}}$ (the dimension of $\mathbf{F}_{\text{lm}}$ is five).

Table 2: Mean accuracy and standard deviation of Resnet50 and VGG16 with KNN or SVM as lesion-level classifier

Method	VGG16		Resnet-50	
	KNN	SVM	KNN	SVM
Accuracy	83.11±2.47	82.66±2.02	87.64±0.26	87.64±0.26

3.5. Effectiveness of the pre-trained model

To show the effectiveness of the pre-trained model, we conducted one experiment for performing optimization from scratch. As illustrated in Fig. 7,

the orange and red lines represent the training loss and validation loss without pre-trained, respectively; the blue and green lines represent the training loss and validation loss with pre-trained, respectively. We observed that by using the pre-trained model, the convergence of our model is faster and can also improve the Lesion-level classification result (Table 3). The supporting assumption is that the large-scale dataset (e.g. ImageNet) can distill the low-level characteristics (e.g. gray-level values and edge filters), which are common for vision tasks [26, 27, 28]. Hence, the performance of our model was improved by using the pre-trained model.

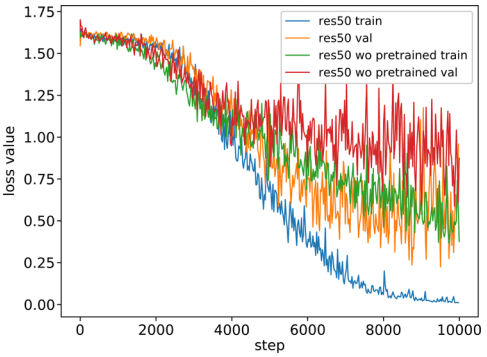


Figure 7: Processes of loss value change. The green and blue lines represent the model with the pre-training, and the red and orange lines represent the model without pre-training.

Table 3: The results of our model with and without pre-training

Method	Proposed method (w/o pre-trained)	Proposed method
Accuracy	70.85±3.92	87.64±0.26

4. Results and discussion

In this section, (1) we will show the mean precision, recall, and the two confusion matrixes of the two test datasets. (2) We will validate the effectiveness of the geometry feature and BD-LSTM block, which are used to extract enhancement patterns. (3) We will also validate the effectiveness of the joint loss function and (4) the effectiveness of combining local and global information. (5) We will also compare the state-of-the-art methods.

Table 4: The configuration of different models

Method	Geometry Information	BD -LSTM Block	Joint Loss	Global Branch Block (GBB)	Local Branch Block (LBB)
Model-A	–	√	√	√	√
Model-B	√	–	√	√	√
Model-C	√	√	–	√	√
Model-D	√	√	√	–	√
Model-E	√	√	√	√	–
Model-F (proposed method)	√	√	√	√	√

In Section 3, we showed the result of different base networks and the effectiveness of the pre-trained model. In this section, for ease of denotation, we use our model to represent our proposed model with ResNet50, pre-trained model and SVM classifier. Furthermore, we used Model-A to Model-F to represent our proposed model with different configurations. Table 4 shows the configuration of different models.

4.1. Classification accuracy

In this subsection, we list the two confusion matrixes together with the precision and recall of the two test datasets in Table 5 and Table 6. We can observe the difficulty in distinguishing between HCC and METS. On the basis of the doctor's clinical diagnosis, they will perform diagnoses by using the patient's medical history and whether multiple tumors exist. However, the input of our model only contains one tumor, thus limiting the receptive field. Hence the undesirable consequence between the HCC and METS is considerable. Furthermore, the ring enhancement, which is one of the most important characteristics of METS, is sometimes unclear. Fig. 8(a) illustrates one misclassified METS sample, which has unclear ring enhancement. Hence, it is misclassified as HCC. Moreover, Fig. 8(b) illustrates another misclassified METS sample, namely, a multi-tumors sample. Given that the input of our model only involves one tumor, it will lead to limited receptive fields (i.e. one tumor cannot recognize another tumor). Therefore, the sample is misclassified. As shown in Table 5 and Table 6, some FNHs are misclassified as HEM because of the non-typical appearance of FNH.

Table 5: Confusion matrix for test dataset1 (Model-F)

Type	Cyst	FNH	HCC	HEM	METS	Recall (%)
Cyst	26	0	0	0	0	100.0
FNH	0	17	0	1	0	95.44
HCC	0	3	19	0	4	73.07
HEM	0	0	3	23	0	88.46
METS	1	0	3	0	24	85.71
Precision (%)	96.29	85.00	76.00	95.83	85.71	–

Table 6: Confusion matrix for test dataset2 (Model-F)

Type	Cyst	FNH	HCC	HEM	METS	Recall (%)
Cyst	17	0	0	0	0	100.0
FNH	0	20	0	2	1	86.95
HCC	0	1	29	4	2	80.55
HEM	0	0	0	17	0	100.0
METS	2	0	2	0	14	77.78
Precision (%)	89.43	95.23	93.54	73.91	82.35	–

4.2. Effectiveness of some proposed blocks

In this subsection, we conducted two group experiments to validate the effectiveness of geometry information and BD-LSTM. As shown in Table 4, Model-A and Model-B represent the model without geometry information and BD-LSTM, respectively. The mean accuracy and standard deviation are shown in Table 7. The second column shows the result without the geometry information. To validate the significance of geometry information, we performed the Wilcoxon's signed-rank test. The p-value is smaller than 0.05, therefore, the use of geometry information can significantly improve the performance of FLL classification. The third column shows the result without a BD-LSTM block. Note that, we replaced the BD-LSTM block with the concatenation operation in Model-B. We can observe that all proposed blocks can improve the performance.

4.3. Effectiveness of the joint loss function

In this subsection, we validated the effectiveness of joint loss function. We compared our model with the Model-C without joint loss, which only uses

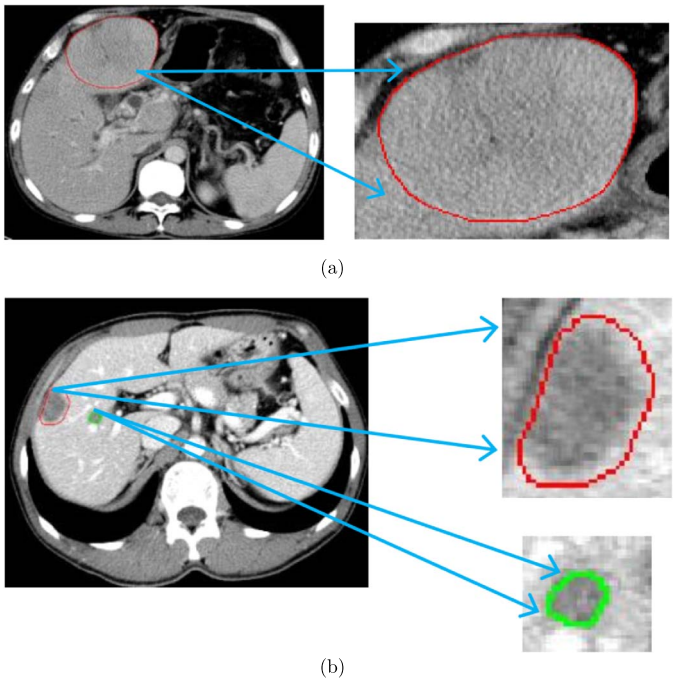


Figure 8: The misclassified samples. (a) is METS, it is misclassified to HCC; (b) the both tumors are METS, they are misclassified to HCC.

$\mathcal{L}_{last}$ to optimize the model. Table 8 shows the result. As mentioned previously, HCC and METS including FNH and HEM, are hard to discriminate. Hence, there is a trade-off between them where one accuracy value is higher; another accuracy value is lower. In Table 8, although the accuracy of HEM and METS in our model is lower, we can observe that they can achieve a better trade-balance point. The higher total accuracy of our model is also another proof of performance improvement.

4.4. Effectiveness of combining local and global information

In this subsection, we will prove that combing global and local information can achieve better performance than utilizing global or local information separately. We conducted two group comparative experiments: one does not have local information (Model-E), and the other does not have global information (Model-D). The results are shown in the Table 9. We can observe that using global information only can achieve better performance than using local information only. It illustrates that global information is more

Table 7: The mean accuracy and standard deviation of the proposed method (Model-F) compared with those of the models without geometry information (Model-A) and BD-LSTM (Model-B)

Method	Model-A	Model-B	**Model-F**
Accuracy	83.32±1.34	82.302±3.27	**87.642±0.26**

Table 8: The mean accuracy and standard deviation of the proposed Model-F compared with those of Model-C (without joint-loss)

Method	Cyst	FNH	HCC	HEM	METS	Total Accuracy
Model-C	**100.0 ±0.0**	46.13 ±1.69	70.19 ±4.81	**96.15 ±3.85**	**84.03 ±1.68**	79.55 ±0.28
Model-F (proposed method)	**100.0 ±0.0**	**90.69 ±3.74**	**76.81 ±3.74**	94.23 ±5.77	81.75 ±3.96	**87.64 ±0.26**

important than local information. The receptive field of the global branch is larger than the local branch, thus leading to the former to achieve better performance intuitively.

We also observed that combining local and global information can achieve better performance than both. It illustrates that there is complementary information that can improve performance between the local and global information. For example, the enhancement boundary, which is very small, is important to diagnose (Fig. 9). However, the important information may be lost during the pooling operation because we only use the lesion input. Hence, the patch is information that complements the local information.

Table 9: The mean accuracy and standard deviation of the proposed Model-F compared with these of Model-E (without local branch) and Model-D (without global branch)

Method	Cyst	FNH	HCC	HEM	METS	Total Accuracy
Model-E	**100.0 ±0.0**	74.67 ±8.00	78.73 ±1.81	92.30 ±7.69	57.14 ±7.14	80.50 ±1.47
Model-D	93.20 ±0.90	86.71 ±8.93	**82.60 ±2.05**	24.2 ±6.56	66.46 ±5.75	71.65 ±3.11
Model-F (proposed method)	**100.0 ±0.0**	**90.69 ±3.74**	76.81 ±3.74	**94.23 ±5.77**	**81.75 ±3.96**	**87.64 ±0.26**

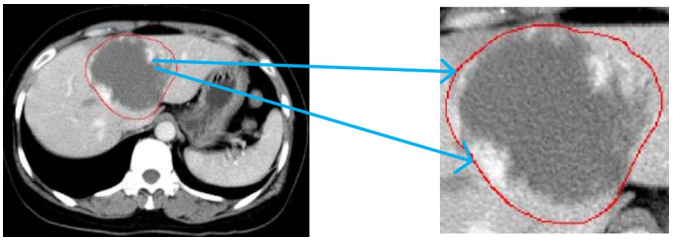


Figure 9: HEM Tumor. It illustrates the importance of combining local and global information.

4.5. Comparison of state-of-the-art methods

To validate the effectiveness of our proposed methods, we compared our results with state-of-the-art methods with low-level features [2], mid-level features (i.e. BoVW-based methods) [6, 7, 8, 9, 10] and high-level features (e.g. CNN-based methods) [12, 13, 28]. For Frid-Adar et al. [12], we also used ResNet50 and multi-phase CT images as input for equality, which can be seen as a local branch. For Wang et al. [28], we used the fine-tuning ResNet50 to classify the lesions. We listed the comparative experimental results in Table 10, which shows that the proposed method (model-F) outperformed the state-of-the-art methods. The proposed method achieves a new baseline for liver tumor classification in multi-phase CT image.

5. Conclusions

In this paper, we proposed a method that uses a combination of convolution neural networks with local and global pathways and bi-directional long short-term memory (CasCRNN-GL-Net), to tackle the classification of focal liver lesions in the multi-phase CT images. The CNN-GLNet extracted the most representative features from each single-phase CT image, and the BD-LSTM helped in extracting the enhancement patterns in multi-phase CT images. Furthermore, we designed a joint loss function that combines the inter-class loss and intra-class loss for performance improvement. We also performed extensive experiments to validate the effectiveness of the proposed framework and blocks. Furthermore, we investigated the effectiveness of combining local and global information. The hidden local and global information is complementary. Hence, combining them can improve performance. We can use any state-of-the-art deep CNNs as base networks for the

Table 10: Comparison to the state-of-the-art method, classification accuracy is represented as mean and standard deviation

Method		Cyst	FNH	HCC	HEM	METS	Total Accuracy
Hand-crafted features	Roy et al. [2]	81.37 ±1.37	71.31 ±1.75	48.10 ±11.90	44.16 ±2.50	38.33 ±11.66	57.19 ±3.16
Mid-level feature (BoVW and its variant	Xu et al. [6]	97.91 ±2.08	73.47 ±15.41	**86.89 ±5.41**	50.69 ±14.69	7.14 ±7.14	66.89 ±2.67
	Xu et al. [9]	78.04 ±1.12	61.46 ±16.30	73.92 ±18.37	60.84 ±3.15	33.92 ±19.64	64.00 ±8.35
	Yu et al. [7]	75.79 ±4.96	63.79 ±25.09	73.71 ±7.05	73.90 ±4.67	42.85 ±14.28	65.05 ±2.67
	Diamant et al. [8]	73.07 ±0.0	74.99 ±2.77	84.61 ±0.0	70.83 ±0.0	37.49 ±1.78	67.94 ±0.82
	Wang et al. [10]	**100.0 ±0.0**	52.95 ±30.37	69.86 ±3.20	65.10 ±7.96	39.28 ±3.57	64.94 ±6.83
High-level feature (CNN based)	Frid-Adar et al. [12]	93.20 ±0.90	86.71 ±8.93	82.60 ±2.05	24.2 ±6.56	66.46 ±5.75	71.65 ±3.11
	Yasaka et al. [13]	95.13 ±1.02	69.92 ±13.40	73.71 ±7.05	62.32 ±20.02	72.22 ±22.22	72.50 ±3.15
	Wang et al. [28]	**100.0 ±0.0**	90.09 ±1.21	85.89 ±2.56	67.19 ±21.04	**82.85 ±9.32**	80.33 ±6.14
Model-F (proposed method)		**100.0 ±0.0**	**90.69 ±3.74**	76.81 ±3.74	**94.23 ±5.77**	81.75 ±3.96	**87.64 ±0.26**

proposed framework. We confirmed the effectiveness of our framework by using both ResNet50 and VGG16 as base networks. The use of ResNet50 as the base network can achieve better performance than VGG16. By using the pre-trained model, our model can converge faster. Moreover, compared with the state-of-the-art methods, our approach improves the baseline classification of focal liver lesions.

For our future work, we plan to implement the proposed model in an end-to-end manner to achieve better performance. Furthermore, building a large-scale liver lesion dataset is still a challenging task. Moreover, the slice-level input may be essential for further improving the result because it contains a larger receptive field. We believe that our proposed framework can be applied to other contrast-enhanced multi-phase CT images.

Acknowledgements

This work was supported by the Major Scientific Project of Zhejiang Lab (No. 2018DG0ZX01) and in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 18H03267 and No. 17H00754.

References

[1] A. B. Ryerson, C. R. Eheman, S. F. Altekruse, J. W. Ward, A. Jemal, R. L. Sherman, S. J Henley, D. Holtzman, A. Lake, A.-M. Noone et al., *Annual Report to the Nation on the Status of Cancer, 1975–2012, featuring the increasing incidence of liver cancer*. Cancer, **122**(9):1312–1337, 2016.

[2] S. Roy, Y. Chi, J. Liu, S. K. Venkatesh and M. S. Brown, *Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions*. IEEE Transactions on Biomedical Engineering, **61**(11): 2768–2778, 2014.

[3] J. Chen, L. Yang, Y. Zhang, M. Alber and D. Z. Chen, *Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation*. Advances in Neural Information Processing Systems, 3036–3044, 2016.

[4] G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray, *Visual categorization with bags of keypoints*. In: Workshop on Statistical Learning in Computer Vision, ECCV, **1**(1-22): 1–2, 2004.

[5] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng and W. Chen, *Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced CT images*. Journal of Digital Imaging, **25**(6): 708–719, 2012.

[6] Y. Xu, L. Lin, H. Hu, D. Wang, W. Zhu, J. Wang, X.-H. Han and Y.-W. Chen, *Texture-specific bag of visual words model and spatial cone matching-based method for the retrieval of focal liver lesions using multiphase contrast-enhanced CT images*. International Journal of Computer Assisted Radiology and Surgery, **13**(1):151–164, 2018.

[7] M. Yu, Q. Feng, W. Yang, Y. Gao and W. Chen, *Extraction of lesion-partitioned features and retrieval of contrast-enhanced liver images*. Computational and Mathematical Methods in Medicine, **2012**, 2012.

[8] I. Diamant, A. Hoogi, C. F. Beaulieu, M. Safdari, E. Klang, M. Amitai, H. Greenspan and D. L. Rubin, *Improved patch-based automated liver lesion classification by separate analysis of the interior and boundary regions*. IEEE Journal of Biomedical and Health Informatics, **20**(6):1585–1594, 2015.

[9] Y. Xu, L. Lin, H. Hu, D. Wang, Y. Liu, J. Wang, X. Han and Y.-W. Chen, *Bag of temporal co-occurrence words for retrieval of focal liver le-*

sions using 3D multiphase contrast-enhanced CT images. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pages 2282–2287, 2016.

[10] J. Wang, X.-H. Han, Y. Xu, L. Lin, H. Hu, C. Jin and Y.-W. Chen, *Sparse codebook model of local structures for retrieval of focal liver lesions using multiphase medical images*. International Journal of Biomedical Imaging, **2017**, 2017.

[11] J. Wang, J. L, X.-H. Han, L. Lin, H. Hu, Y. Xu, Q. Chen, Y. Iwamoto and Y.-W. Chen, *Tensor-based sparse representations of multi-phase medical images for classification of focal liver lesions*. Pattern Recognition Letter, **130**:207–215, 2020. MR3526603

[12] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger and H. Greenspan, *Modeling the intra-class variability for liver lesion detection using a multi-class patch-based CNN*. In: International Workshop on Patch-based Techniques in Medical Imaging, pages 129–137, 2017.

[13] K. Yasaka, H. Akai, O. Abe and S. Kiryu, *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.

[14] D. Liang, L. Lin, H. Hu, Q. Zhang, Q. Chen, X. Han, Y.-W. Chen et al., *Combining convolutional and recurrent neural networks for classification of focal liver lesions in multi-phase CT images*. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 666–675, 2018.

[15] C. Dong, Y.-W. Chen, L. Lin, H. Hu, C. Jin, H. Yu, X.-H. Han and T. Tateyama, *Simultaneous segmentation of multiple organs using random walks*. Journal of Information Processing, **24**(2):320–329, 2016.

[16] C. Dong, Y.-W. Chen, T. Seki, R. Inoguchi, C.-L. Lin and X.-H. Han, *Non-rigid image registration with anatomical structure constraint for assessing locoregional therapy of hepatocellular carcinoma*. Computerized Medical Imaging and Graphics, **45**:75–83, 2015.

[17] D. Liang, L. Lin, H. Hu, Q. Zhang, Q. Chen, X. Han, Y.-W. Chen et al., *Residual convolutional neural networks with global and local pathways for classification of focal liver lesions*. In: Pacific Rim International Conference on Artificial Intelligence, pages 617–628, 2018.

[18] Q. Dou, C. Ouyang, C. Chen, H. Chen and P.-A. Heng, *Unsupervised cross-modality domain adaptation of convnets for biomedical image*

segmentations with adversarial loss. arXiv preprint arXiv:1804.10916, 2018.

[19] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, *How transferable are features in deep neural networks?* In: Advances in Neural Information Processing Systems, pages 3320–3328, 2014.

[20] F. A. Gers, J. Schmidhuber and F. Cummins, *Learning to forget: Continual prediction with LSTM*, 1999.

[21] Y. Wen, K. Zhang, Z. Li and Y. Qiao, *A discriminative feature learning approach for deep face recognition*. In: European Conference on Computer Vision, pages 209–304, 2016.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and F.-F. Li, *Imagenet: A large-scale hierarchical image database*. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, 2009.

[23] P. Urban, X. Défago and A. Schiper, *Neko: A single environment to simulate and prototype distributed algorithms*. In: Proceedings 15th International Conference on Information Networking, pages 503–511, 2001.

[24] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

[25] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014.

[26] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway and J. Liang, *Convolutional neural networks for medical image analysis: Full training or fine tuning?* IEEE Transactions on Medical Imaging, **35**(5):1299–1312, 2016.

[27] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin et al., *Interactive medical image segmentation using deep learning with image-specific fine tuning*. IEEE Transactions on Medical Imaging, **37**(7):1562–1573, 2018.

[28] W. Wang, Y. Iwamoto, X. Han, Y.-W Chen, Q. Chen, D. Liang, L. Lin, H. Hu and Q. Zhang, *Classification of focal liver lesions using deep learning with fine-tuning*. In: Proceedings of the 2018 International Conference on Digital Medicine and Image Processing, pages 56–60, 2018.

Dong Liang
College of Computer Science and Technology
Zhejiang University
Hangzhou
China
E-mail address: cs_liangdong@zju.edu.cn

Yingying Xu
College of Computer Science and Technology
Zhejiang University
Hangzhou
China
E-mail address: cs_ying@zju.edu.cn

Lanfen Lin
College of Computer Science and Technology
Zhejiang University
Hangzhou
China
E-mail address: llf@zju.edu.cn

Nan Zhou
College of Computer Science and Technology
Zhejiang University
Hangzhou
China
E-mail address: zhounan_1211@163.com

Hongjie Hu
Department of Radiology
Sir Run Run Shaw Hospital
Hangzhou
China
E-mail address: hongjiehu@zju.edu.cn

Qiaowei Zhang
Department of Radiology
Sir Run Run Shaw Hospital
Hangzhou
China
E-mail address: radiologist@163.com

Qingqing Chen
Department of Radiology
Sir Run Run Shaw Hospital
Hangzhou
China
E-mail address: qingqingchen@zju.edu.cn

Xianhua Han
College of Information Science and Engineering
Ritsumeikan University
Kyoto
Japan
E-mail address: hanxhua@yamaguchi-u.ac.jp

Yutaro Iwamoto
College of Information Science and Engineering
Ritsumeikan University
Kyoto
Japan
E-mail address: yiwamoto@fc.ritsumei.ac.jp

Yen-Wei Chen
College of Information Science and Engineering
Ritsumeikan University
Kyoto
Japan
E-mail address: chen@is.ritsumei.ac.jp

Received August 3, 2020