Combining Convolutional and Recurrent Neural Networks for Classification of Focal Liver Lesions in Multi-phase CT Images

Dong Liang[1], Lanfen Lin[1(✉)], Hongjie Hu[2], Qiaowei Zhang[2], Qingqing Chen[2], Yutaro lwamoto[3], Xianhua Han[3], and Yen-Wei Chen[1,3]

[1] Department of Computer Science and Technology, Zhejiang University, Hangzhou, China
llf@zju.edu.cn
[2] Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou, China
[3] College of Information Science and Engineering, Ritsumeikan University, Kusatsu, Shiga, Japan

Abstract. Computer-aided diagnosis (CAD) systems are useful for assisting radiologists with clinical diagnoses by classifying focal liver lesions (FLLs) based on multi-phase computed tomography (CT) images. Although many studies have conducted in the field, there still remain two challenges. First, the temporal enhancement pattern is hard to represent effectively. Second, the local and global information of lesions both are necessary for this task. In this paper, we proposed a framework based on deep learning, called ResGL-BDLSTM, which combines a residual deep neural network (ResNet) with global and local pathways (ResGL Net) with a bi-directional long short-term memory (BD-LSTM) model for the task of focal liver lesions classification in multi-phase CT images. In addition, we proposed a novel loss function to train the proposed framework. The loss function is composed of an inter-loss and intra-loss, which can improve the robustness of the framework. The proposed framework outperforms state-of-the-art approaches by achieving a 90.93% mean accuracy.

Keywords: Deep learning · ResGLNet · BD-LSTM
Liver lesions classification · Computer-aid diagnosis (CAD) system

1 Introduction

Liver cancer is the second most common cause of cancer-related deaths worldwide among men, and the sixth among women [1]. Radiological examinations, such as computed tomography (CT) images and magnetic resonance images (MRI) are the primary methods of detecting liver tumors. Computer-aided diagnosis (CAD) systems play an important role in the early and accurate detection and classification of FLLs.

Currently, multi-phase CT images, which are also known as dynamic CT images, are widely used to detect, locate and diagnose focal liver lesions. Multi-phase CT scans are generally divided into four phases (i.e. non-contrast phase, arterial phase, portal phase, delay phase). Between the non-contrast phase and the delay phase, the vascularity and

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11071, pp. 666–675, 2018.
https://doi.org/10.1007/978-3-030-00934-2_74

the contrast agent enhancement patterns of the liver masses can be assessed. We observe that, when human experts diagnose the type of FLLs, they tend to zoom out the CT images to figure out the detail of lesions [2], and they also need to look back or forward in different phases. The observation interprets the importance of the combination of local with global information and the temporal enhancement pattern.

Some published studies have reported on the characterization of FLLs using multiphase images to capture the temporal information among phases. Roy et al. [3] proposed a framework to extract spatiotemporal features from multiphase CT volumes for the characterization of FLLs. In addition to conventional density features (the normalized average intensity of a lesion) and texture features (the gray-level co-occurrence matrix [GLCM]), temporal density and texture features (the intensity and texture enhancement over the three enhancement phases compared with the non-contrast phase), were employed. Compared with low-level features, the mid-level features such as bag-of-visual-words (BoVW) and its variants have proven to be considerably more effective for classifying FLLs [4–9]. In most of the BoVW-based methods, the histograms in each phase are separately extracted and then they are concatenated as a spatiotemporal feature [5, 8, 9] or the averaged histogram over multiple phases is used to represent the multi-phase images [4]. They ignore the temporal enhancement information and relationship among phases.

In recent years, the high-level feature representation of deep convolutional neural networks (DCNN) has proven to be superior to hand-crafted low-level features and mid-level features [10]. Deep learning techniques have also been applied to medical image analysis and computer-aided detection and diagnosis. However, there have been very few studies on the classification of focal liver lesions. Frid-Arar et al. [11] proposed a multi-scale patch-based classification framework to detect focal liver lesions. Yasaka et al. [12] proposed a convolutional neural network with three channels corresponding to three phases (NC, ART and DL) for the classification of liver tumors in dynamic contrast-enhanced CT images. The method can extract high-level temporal and spatial features, resulting in a higher classification accuracy compared with the state-of-the-art methods. The limitation is that it lacks information on image pattern enhancements.

In this paper, we propose a framework based on deep learning, called ResGL-BD-LSTM, which combines a residual network (ResNet) with global and local pathways (ResGLNet) [13] and a bi-directional long short-term memory (BD-LSTM) model for the classification of focal liver lesion. The main contributions are summarized as follows:

(1) We extract features from each single phase CT image via the ResGLNet. The input of the ResGLNet is a pair (patch and ROI) that represent the local and global information, respectively, to handle inter-class similarities.
(2) We extract an enhancement pattern, hidden in multi-phase CT images, via the 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.

(3) We propose a new 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. The inter-loss minimizes the inter-class variations and the intra-loss minimizes the intra-class variations, updating the center value using a back-propagation process.

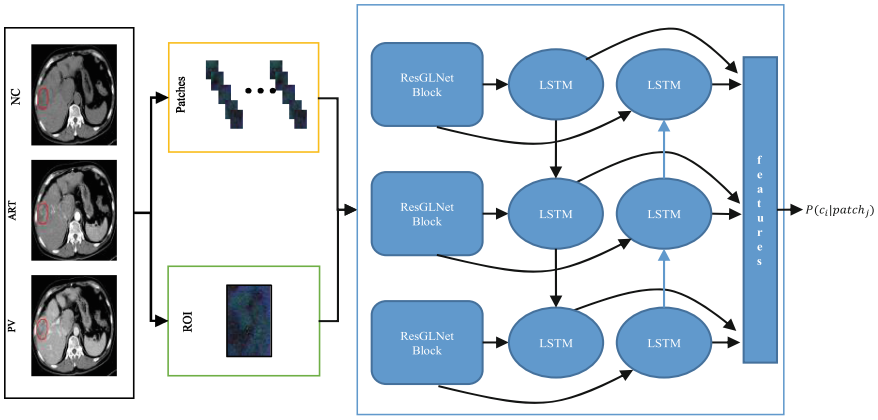

Fig. 1. The flowchart of our framework

2 Methodology

A flowchart of the proposed framework is shown in Fig. 1. The ResGLNet block, which extracts local and global information from each single phase, will be described in detail in Sect. 2.1. The BD-LSTM block, which extracts the enhancement pattern, will be described in detail in Sect. 2.2. The method combining the ResGLNet block and BD-LSTM block will be described in Sect. 2.3. We will introduce the loss function and training strategy of the framework in Sect. 2.4. In Sect. 2.5, we describe the features extracted from the label map, and how we accomplish the lesion-based classification.

2.1 ResGLNet

In this sub-section, we describe ResGLNet block, which was proposed in our previous work [13]. The ResGLNet involves a local pathway and global pathway. Intuitively, these extract local and global information, respectively. The employed ResGLNet is an extension of the ResNet proposed by [10]. We utilize three ResGLNet blocks, which each have the same architecture but do not share weights with each other, to extract the information of the three respective phases. In each ResNet block, we used 19 convolutional layers, one pooling layer (avg-pooling), and one fully connected layer. Each convolution layer was followed by a rectified linear unit (ReLU) activation function and a batch normalization layer.

Global Pathway. First, we apply a random walk-based interactive segmentation algorithm [14] to segment healthy tissue and focal liver lesions. The segmented results were checked by two experienced radiologists. The segmentation was performed for each phase image separately. During a clinical CT study, the spatial placement of tissues formed in multiple phases exhibits some aberration, owing to differences in a patient's body position, respiratory movements, and the heartbeat. Therefore, to obtain a factual variation of the density over phases, a non-rigid registration technique in order to localize a reference lesion in other phases [15]. Each segmented lesion image (i.e., 2D slice image) was resized to 128×128. The resized images were then used as input for global pathway training and testing.

Local Pathway. Patches were extracted from ROIs. Each patch has a label, $c \in \{c_0, c_1, c_2, c_3\}$ where c_0 represents a cyst, c_1 represents an focal nodular hyperplasia (FNH), c_2 represents an hepatocellular carcinoma (HCC) and c_3 represents an hemangioma (HEM). Owing to the different lesions varying significantly in size, extreme imbalances occur among the patch categories. To solve this problem, the pace value is derived in Eq. (1):

$$pace_i = \begin{cases} floor\left(\sqrt{\frac{w_i * h_i}{\epsilon}}\right), & w_i * h_i > \epsilon \\ 1, & w_i * h_i \leq \epsilon \end{cases} \tag{1}$$

where i represents the i-th ROI; $pace_i$ is the pace of i-th ROI for extracting the patches, w_i and h_i respectively represent the width and height of the i-th ROI, ϵ represents a threshold that can limit the number of patches, and the floor function represents rounding-down. For the testing dataset, we still set the pace to 1. As in the global pathway approach, we resized the patches to 64×64.

2.2 BD-LSTM

A recurrent neural network (RNN) can maintain self-connected status acting as a memory to remember previous information when it processes sequential data. Long-short term memory (LSTM) 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. 1, work in two opposite directions to extract useful information from sequential data. The enhancement information carried in the two layers of LSTM is concatenated as the output. One layer is in the z^--direction, and extracts the enhancement pattern from the NC phase through the PV phase and the other is in the z^+-direction and extracts the anti-enhancement pattern from the PV phase through the NC phase.

2.3 Combining ResGLNet and BD-LSTM

The motivation of performing focal liver lesions classification based on multi-phases CT images by combining ResGLNet and BD-LSTM is to employ multi-phases CT images as sequential data. The ResGLNet extracts the information (i.e., intra-phase information) based on a single phase. The BD-LSTM distills enhancement information (i.e., inter-phase information) among three phases, and the length of sequential data is a constant number (i.e., 3). The two blocks work in coordination, as follows.

The output of the three ResGLNet blocks, as a sequential data, constitutes the input of the BD-LSTM. Furthermore, the output of the two layers LSTM (i.e., BD-LSTM), representing patches, constitutes the input of the fully connected layer. The softmax layer following the last fully connected layer produces output that gives the result of the patch-based classification.

2.4 Training Strategy

Loss Function. Let N be the batch size and ω^t be the weights in the t-th (t = 1, 2, … T) layer. We use $\boldsymbol{W}$ to denote the weights of the mainstream network (involving three ResGLNet blocks and a BD-LSTM block). We used $\widehat{\mathrm{W}}_{\mathrm{local}}$ and $\widehat{\mathrm{W}}_{\mathrm{global}}$ to represent the weights of the local and global pathways (involving three ResGLNet blocks, the same below), respectively. Furthermore, $p(j \mid x_i; \boldsymbol{W})$ represents the probability of the i-th patch belonging to the j-th class. We define $p\left(j \middle| x_i; \widehat{\mathrm{W}}_{\mathrm{global}}\right)$ and $p\left(j \middle| x_i; \widehat{\mathrm{W}}_{\mathrm{local}}\right)$ similarly. The definitions of cross-entropy are as follows:

$$\mathcal{L}_{last} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{K} -p(j \mid x_i; \boldsymbol{W}) * \log(p(j \mid x_i; \boldsymbol{W})) \tag{2}$$

Thus, we can obtain the definition of $\mathcal{L}_{local}$ and $\mathcal{L}_{global}$ for the same reason. And the definition of the inter-loss that as follows:

$$\mathcal{L}_{inter} = \frac{1}{2} * \mathcal{L}_{last} + \frac{1}{4} * \mathcal{L}_{local} + \frac{1}{4} * \mathcal{L}_{global} \tag{3}$$

The definition of the intra-loss (i.e. center loss [16]) is as follows:

$$\mathcal{L}_{intra} = \frac{1}{2}\sum_{i=1}^{N} \left\| f_i - c_{y_i} \right\|^2 \tag{4}$$

Here $\boldsymbol{f_i}$ 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 process of back-propagation. To accelerate our training, we conduct the update operation based on each batch, instead of basing it on the entire training set. Note that, in this case,

some of the centers may not change. The method employed to update the centers is described as follows:

$$\begin{aligned} \boldsymbol{c}_j &= \boldsymbol{c}_j - \Delta \boldsymbol{c}_j = \boldsymbol{c}_j - \frac{\partial \mathcal{L}}{\partial \boldsymbol{c}_j} \\ &= \boldsymbol{c}_j - \alpha * \frac{\sum_{i=1}^{N} \delta(y_i == j)(\boldsymbol{f}_i - \boldsymbol{c}_j)}{1 + \sum_{i=1}^{N} \delta(y_i == j)} \end{aligned} \tag{5}$$

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

Finally, we adopt a joint loss that combines the intra-loss and inter-loss to train the frameworks. The formulation of the optimized loss is given in Eq. 6.

$$\mathcal{L} = \mathcal{L}_{inter} + \lambda \mathcal{L}_{intra} \tag{6}$$

Training Process. Our framework is split into two phases. The first is the training phase, and the second is the 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 stop. We aggregated the label maps belonging to the training and validation datasets after the model was trained. Next, we used the training and validation label maps as input to the support vector machine (SVM) classifier. Then, we also determined the parameters of the SVM (classifier of lesions) using the effectiveness of label map of the validation dataset.

2.5 Post-processing of Label Map and Classification of Lesions

After training, we aggregated the label map of each lesion. Then, we extracted features from the label map. The features are as follows:

$$feature_i = \{\beta_{i0}, \beta_{i1}, \beta_{i2}, \beta_{i3}\} \tag{7}$$

Here $feature_i$ represents the feature vector of i-th label map, and β_{ij}, is derived in Eq. (8), denotes the proportion of pixels belonging to the j-th category of in the i-th label map. Then we use the SVM to achieve lesion-based classification.

$$\beta_{ij} = \frac{the\ number\ of\ pixels\ belong\ to\ jth\ category}{the\ total\ pixels\ in\ ith\ label\ map} \tag{8}$$

3 Experiments

3.1 Data and Implementation

A total of 480 CT liver slice images were used, containing four types of lesions confirmed by pathologists, (i.e., Cyst, HEM, FNH, and HCC). The distribution of our dataset is shown in Table 1. 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 $0.57 - 0.89$. In our experiment, we randomly split our dataset into a training dataset, a validation dataset, and a testing dataset. In order to eliminate the effect of randomness, we conduct the partition operation twice, and form two groups of dataset.

Table 1. The distribution of database.

Type	Cyst		FNH		HCC		HEM	
	Set1	Set2	Set1	Set2	Set1	Set2	Set1	Set2
Training	61	69	71	60	75	69	62	79
Validation	23	17	25	23	31	36	36	17
Testing	26	24	18	31	26	27	26	28
Total	110		114		132		124	

Our framework was implemented using the Tensorflow library. We initialized the parameters via the Gaussian distribution. We used a momentum optimizer to update our parameters by setting the learning rate initialized as 0.01 and the momentum coefficient to 0.9. We set the batch size as 100. The parameters for our algorithm were $\lambda = 0.1$, $\alpha = 0.2$, $\epsilon = 128$, and *patch size* = 7.

3.2 Results

In order to validate the effectiveness of or proposed methods. We compared our results with the state-of-art methods with low-level features [2], mid-level features [4–8] and CNN with local information [10] and global information [11]. We also compared our proposed methods with different architectures: ResNet with local patch (w/o intra-loss), ResGLNet [13], ResGL-BDLSTM (w/o intra-loss), and ResGL-BDLSTM (with intra-loss). The comparison results (classification accuracy) are summarized in Table 2. It can be seen that our proposed methods outperformed the state-of-the-art methods [3, 5–9, 11, 12]. The ResNet with local and global pathways outperformed the ResNet with local patch only. The classification accuracy was significantly improved by adding the BD-LSTM model, as well as the intra-loss.

Table 2. Comparison results (classification accuracy (%) is represented as mean and standard deviation)

Method	Cyst	FNH	HCC	HEM	Total Accuracy
Roy et al. [3]	97.81 ± 3.1	77.27 ± 6.43	58.83 ± 2.39	56.41 ± 14.5	71.84 ± 0.04
Yang et al. [5]	88.30 ± 10.6	74.64 ± 28.1	75.50 ± 2.0	81.32 ± 6.2	78.81 ± 3.4
Wang et al. [8]	85.90 ± 3.6	65.14 ± 32.8	83.12 ± 7.5	67.99 ± 20.0	74.06 ± 5.7
Xu et al. [9]	68.75 ± 2.9	73.53 ± 4.1	87.25 ± 1.3	76.92 ± 10.8	77.04 ± 3.1
Diamant et al. [6]	82.21 ± 7.4	70.00 ± 18.8	85.04 ± 10.2	76.90 ± 20.3	77.82 ± 1.2
Xu et al. [7]	92.15 ± 5.21	69.08 ± 20.1	85.04 ± 10.2	84.31 ± 0.4	82.11 ± 7.4
Frid-Adar et al. [11] (CNN with local)	100.0 ± 0.0	78.20 ± 0.5	84.37 ± 16.6	40.67 ± 16.2	76.16 ± 0.6
Yasaka et al. [12] (CNN with global)	97.92 ± 2.9	82.26 ± 25.1	86.82 ± 2.32	85.16 ± 0.7	87.26 ± 7.7
ResNet_Local	100.0 ± 0.0	71.27 ± 6.5	80.89 ± 11.18	85.41 ± 8.8	84.12 ± 6.1
ResGLNet [13]	97.92 ± 2.9	81.99 ± 5.9	85.11 ± 15.6	85.42 ± 2.9	88.05 ± 4.8
ResGL-BDLSTM (without intra-loss)	98.08 ± 2.2	90.19 ± 8.9	88.74 ± 5.0	81.25 ± 8.8	89.77 ± 3.59
ResGL-BDLSTM	100.0 ± 0.0	86.74 ± 4.1	88.82 ± 10.3	87.75 ± 5.5	90.93 ± 0.7

4 Conclusions

In this paper, we proposed a method using combined residual local and global pathways and bi-directional long short-term memory (ResGL-BDLSTM), to tackle the classification of focal liver lesions. The ResGLNet extracts the most representative features from each single phase CT image, and the BD-LSTM helps to extract the enhancement patterns in multi-phases CT images. The experimental results demonstrated that our framework outperforms other state-of-the-art methods. In the future work, we are going to build a large scale liver lesions dataset and to construct an end-to-end framework that achieves lesion-based classification via one model. We believe that our proposed framework can be applied to other contrast-enhanced multi-phases CT images.

Acknowledgements. This work was supported in part by the National Key R&D Program of China under the Grant No. 2017YFB0309800, in part by the Key Science and Technology Innovation Support Program of Hangzhou under the Grant No. 20172011A038, 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. Ryerson, A.B., 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. Chen, J., et al.: Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: Advances in Neural Information Processing Systems (2016)
3. Roy, S., et al.: Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions. IEEE Trans. Biomed. Eng. **61**(11), 2768–2778 (2014)
4. Yu, M., et al.: Extraction of lesion-partitioned features and retrieval of contrast-enhanced liver images. Comput. Math. Meth. Med. **2012**, 12 (2012)
5. Yang, W., et al.: Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single-and multiphase contrast-enhanced CT images. J. Digital Imaging **25** (6), 708–719 (2012)
6. Diamant, I., et al.: Improved patch-based automated liver lesion classification by separate analysis of the interior and boundary regions. IEEE J. Biomed. Health Inform. **20**(6), 1585–1594 (2016)
7. Xu, Y., et al.: Bag of temporal co-occurrence words for retrieval of focal liver lesions using 3D multiphase contrast-enhanced CT images. In: Proceedings of 23rd International Conference on Pattern Recognition (ICPR 2016), pp. 2283–2288 (2016)
8. Wang J., et al.: Sparse codebook model of local structures for retrieval of focal liver lesions using multiphase medical images. Int. J. Biomed. Imaging **2017**, 13 p. (2017)
9. Xu, Y., et al.: 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. Int. J. Comput. Assist. Radiol. Surg. **13**(1), 151–164 (2018)
10. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
11. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Modeling the intra-class variability for liver lesion detection using a multi-class patch-based CNN. In: Wu, G., Munsell, Brent C., Zhan, Y., Bai, W., Sanroma, G., Coupé, P. (eds.) Patch-MI 2017. LNCS, vol. 10530, pp. 129–137. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67434-6_15
12. Yasaka, K., et al.: Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study. Radiology **286**(3), 887–896 (2017)
13. Liang, D., et al.: Residual convolutional neural networks with global and local pathways for classification of focal liver lesions. In: Geng, X., Kang, B.H. (eds.) PRICAI 2018: Trends in Artificial Intelligence. PRICAI 2018. LNCS, vol. 11012, pp. 617–628. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97304-3_47

14. Dong, C., et al.: Simultaneous segmentation of multiple organs using random walks. J. Inf. Process. **24**(2), 320–329 (2016)
15. Dong, C., et al.: Non-rigid image registration with anatomical structure constraint for assessing locoregional therapy of hepatocellular carcinoma. Comput. Med. Imaging Graph. **45**, 75–83 (2015)
16. Wen, Y., Zhang, K., Li, Z., Qiao, Yu.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31