Local and Non-local Deep Feature Fusion for Malignancy Characterization of Hepatocellular Carcinoma

Tianyou Dou[1], Lijuan Zhang[2], Hairong Zheng[2], and Wu Zhou[1(✉)]

[1] School of Medical Information Engineering,
Guangzhou University of Chinese Medicine, Guangzhou, China
`zhouwu@gzucm.edu.cn`

[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences,
Shenzhen, China

Abstract. Deep feature derived from convolutional neural network (CNN) has demonstrated superior ability to characterize the biological aggressiveness of tumors, which is typically based on convolutional operations repeatedly processed within a local neighborhood. Due to the heterogeneity of lesions, such local deep feature may be insufficient to represent the aggressiveness of neoplasm. Inspired by the non-local neural networks in computer vision, the non-local deep feature may be remarkably complementary for lesion characterization. In this work, we propose a local and non-local deep feature fusion model based on common and individual feature analysis by extracting common and individual components of local and non-local deep features to characterize the biological aggressiveness of lesions. Specifically, we first design a non-local subnetwork for non-local deep feature extraction of neoplasm, and subsequently combine local and non-local deep features with a specific designed fusion subnetwork based on common and individual feature analysis. Experimental results of malignancy characterization of clinical hepatocellular carcinoma (HCC) with Contrast-enhanced MR images demonstrate several intriguing features of the proposed local and non-local deep feature fusion model as follows: (1) Non-local deep feature outperforms local deep feature for lesion characterization; (2) The fusion of local and non-local deep feature yields further improved performance of lesion characterization; (3) The fusion method of common and individual feature analysis outperforms the method of simple concatenation and the method of deep correlation model.

1 Introduction

Hepatocellular carcinoma (HCC) is the most common primary hepatic malignancy, ranking second in the world for the cause of death from tumors [1]. Malignancy of HCC is an important prognostic factor that affects recurrence and survival after liver transplantation or surgical resection in clinical practice [2]. MR imaging has played a significant role in the diagnosis of HCC, in which

there are a variety of studies that address the malignancy characterization of HCC by identifying imaging features [3,4]. However, such morphological features are generally dependent on empirical manual design, which are often insufficient to characterize the heterogeneity of the tumor.

Deep features relied on data-driven learning from samples demonstrate superior ability to characterize tumors [5]. Recently, deep feature in the arterial phase of Contrast-enhanced MR has been verified to outperform texture features for malignancy characterization of HCC [6]. Such local deep feature is typically based on convolutional operations repeatedly processed within a local neighborhood. More recently, a non-local neural network has been illustrated for the task of video classification in computer vision, which is based on a non-local operation that allows distant pixels to make contribution to the response at a position as a weighted mean of features from all the distant pixels [7]. We hypothesize that such non-local deep feature may be remarkably applicable and complementary to local deep feature for malignancy characterization of HCC.

More importantly, it is essential to take full advantage of the local and non-local deep features by optimal fusion for lesion characterization. One simple way for fusing information is concatenating deep features [8] or integrating multi-modal results based on weighted summation [9]. Recently, deep correlational model has been proposed to extract maximum correlated representation of deep features from multimodal by canonic correlation analysis for lesion characterization [10]. However, only shared or correlated component of deep features between modals are extracted, neglecting the influence of separation of deep features across modals for characterization. As a matter of fact, a common part to be shared and a modal-specific part from features of the color and depth information have been recovered to represent the implicit relationship between different modalities for RGB-D object recognition [11,12]. We hypothesize that both the correlated component and separated component between local and non-local deep features of neoplasm may play significant roles in malignancy characterization of HCC.

In this work, we propose a local and non-local deep feature fusion model to characterize the malignancy of HCC. The proposed model first extracts local and non-local deep feature of neoplasm separately, and subsequently recovers common and individual components of local and non-local deep features based on common and individual feature analysis. Specifically, the learned common and individual features can reflect the implicit relationship of local and non-local deep features, which further improve the performance of malignancy characterization of HCC.

2 Method

2.1 Local Deep Feature Extraction

The local deep feature extraction consists of multiple repetitions of convolutional layer with activation function. Given the input feature of image x in CNN, the local deep feature y is obtained by $y = \sigma(Wx + b)$, where W is a convolutional

filter based on a convolutional operation that sums up the weighted input in a local neighborhood, b is the bias term, σ is the rectified linear unit (ReLU) active function.

2.2   Non-local Deep Feature Extraction

The non-local deep feature extraction is based on the conventional non-local mean operation defined in deep neural network as follows [7]

$$y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j) \tag{1}$$

where i is the index of a position to be computed and j is the index of all possible positions. x is the input image and y is the output non-local feature of the same size as x. A similarity function f computes a scalar that manifests approximation between i and j. The function g computes a representation of the input image at the position j. The response is normalized by a factor $C(x)$.

In this work, the g is considered in the form of a linear embedding as $g(x_j) = W_g x_j$, where W_g is a weight matrix to be learned. Furthermore, the similarity function f is considered by the embedded Gaussian as $f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)}$, where $\theta(x_i) = W_\theta x_i$, and $\phi(x_j) = W_\phi x_j$ are two embeddings.

We set $C(x) = \sum_{\forall j} f(x_i, x_j)$, and for a given i, $\frac{1}{C(x)} f(x_i, x_j)$ becomes the softmax computation along the dimension j. Therefore, the output non-local deep feature y becomes

$$y = softmax(x^T W_\theta^T W_\phi x) W_g x \tag{2}$$

where W_g, W_θ and W_ϕ are three weight matrices to be learned. Inspired by the work of [7] in video classification, an implementation of the non-local deep feature map y of neoplasm is described in Fig. 1. Different from the work of [7] in video classification, we conduct the non-local operation directly for the non-local deep feature extraction of neoplasm without considering the residual connection.
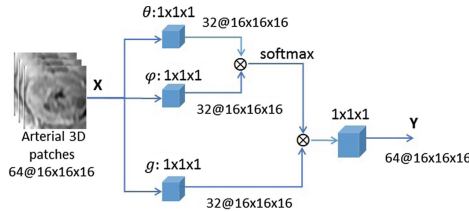
Fig. 1. An implementation of the 3D non-local deep feature map. $\otimes$ denotes matrix multiplication, and "$1 \times 1 \times 1$" denotes $1 \times 1 \times 1$ convolutions. Note that the softmax operation is performed on each row, and we set the number of channels in x to 64.

2.3   Correlation and Individual Feature Analysis

Given two local and non-local deep feature sets $\{Y_i \in R^{(I_i \times J)}, i = 1, 2\}$, the Correlation and individual feature analysis is to extract common and individual components between the two deep feature sets Y_1 and Y_2 in disciplines. Each feature set Y_i is typically decomposed into three terms as follows [13]:

$$Y_i = J_i + A_i + R_i, \quad i = 1, 2 \tag{3}$$

where $J_i \in R^{(I_i \times J)}$ and $A_i \in R^{(I_i \times J)}$ are low-rank matrices, denoting common component between sets and individual component associated with each set, respectively. $R_i \in R^{(I_i \times J)}$ is a matrix denoting residual noise. In order to facilitate the identification of common and individual components, the rows of J and A_i should be mutually orthogonal. Hence, the common component J_i and individual component A_i can be represented by the original deep feature Y_i as

$$J_i = V_i^T V_i Y_i, \quad A_i = Q_i^T Q_i Y_i \tag{4}$$

where V_i is the mapping matrix that projects the original deep feature Y_i into the common component J_i, and Q_i is the mapping matrix that projects the original deep feature Y_i into the individual component A_i. As J_i and A_i should be unrelated and not contaminated by each other, the mapping matrix V_i and Q_i should be orthogonal to each other as $V_i^T Q_i = 0$.

The purpose of extracting the common and individual components between the two local and non-local deep features $\{Y_i \in R^{(I_i \times J)}, i = 1, 2\}$ is solving the constrained least-squares problem:

$$\begin{aligned} min \quad & ||V_1 Y_1 - V_2 Y_2||_F^2, \\ s.t. \quad & Y_i = V_i^T V_i Y_i + Q_i^T Q_i Y_i, i = 1, 2 \\ & V_i^T Q_i = 0, i = 1, 2 \end{aligned} \tag{5}$$

Where $|| \cdot ||_F$ is the Frobenius norm. In this work, alternating optimization is adopted to minimize the constraint least squares problem for all the variable V_i and Q_i. Based on the Lagrange multiplier criterion, the Lagrange function to minimize the constrained least-squares problem is

$$\begin{aligned} \iota(\phi, \theta) = & ||V_1 Y_1 - V_2 Y_2||_F^2 + \sum_{i=1}^{2} \phi_i ||Y_i - V_i^T V_i Y_i - Q_i^T Q_i Y_i||_F^2 \\ & + \sum_{i=1}^{2} \theta_i ||V_i^T Q_i||_F^2 \end{aligned} \tag{6}$$

Where ϕ_i and θ_i are the positive Lagrange multipliers related to the two linear constraints. In this work, we first learn the mapping matrices V_i to map the local and non-local deep features Y_i into the common feature space J_i separately, and then we use Singular Value Decomposition (SVD) to construct the orthogonal basis Q_i of the matrix V_i. Finally, the common component J_i and individual component A_i are obtained by V_i and Q_i according to Eq. (4).

2.4   Local and Nonlocal Deep Feature Fusion Framework

Figure 2 showed the proposed local and non-local deep feature fusion framework. With respect to the extraction of 3D local deep feature by conventional CNN, the convolutional layer was determined by convolving the extracted 3D patches $(16 \times 16 \times 16)$ with a 3D convolution filter $(3 \times 3 \times 3)$ to get the convolution feature maps of the original 3D patch, followed by a pooling layer to perform downsampling operation along the 3D dimensions. In addition, the non-local deep feature can be obtained by the non-local operation as demonstrated in the previous Sect. 2.2. Subsequently, the fusion layer performed the correlation and individual feature analysis to recover common and individual components from the local and non-local deep features. The common component J_1 or J_2 and the individual component A_1 and A_2 are concatenated as the output of local and non-local deep feature fusion, followed by the fully-connected layer and the softmax layer to yield the classification results of low-grade or high-grade of HCC.
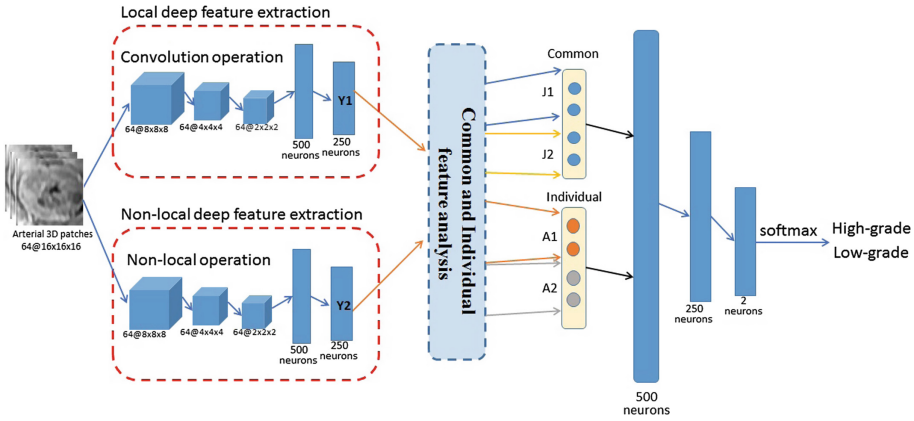

Fig. 2. The proposed local and non-local deep feature fusion framework.

2.5   The Implementation

The proposed framework is implemented by python on the platform of TensorFlow, and the configuration of GPUs used in this work is NVIDIA GeForce GTX1080. The whole network is trained in an end-to-end manner. For the optimization, we use the well-known Adam algorithm [14] for Stochastic Optimization to minimize the objective function. The number of iterations is set to 15000. The initialization of the learning rate is set to 1e–4, and the decay of the learning rate is set to 0.99.

3   Results

The accuracy, sensitivity and specificity are quantitatively computed for malignancy characterization of HCC, and the 4-fold cross-validation with 10 repetitions is adopted to evaluate the performance of the proposed framework.

3.1   Subjects, MR Imaging and Histology Information

Forty-six HCC patients with 46 HCCs are included for this retrospective study from October 2011 to September 2015. Contrast-enhanced MR images with Gd-DTPA agent administration are acquired with a 3.0T MR scanner (Signa Excite HD 3.0T, GE Healthcare, Milwaukee, WI, USA), including pre-contrast, arterial, portal venous, and delayed phase images. The pathological information of HCCs is retrieved from the clinical histology report, including Edmondson grade I (1), II (20), III (24) and IV (1) for these forty-six HCCs. Clinically, Edmondson grade I and II are low-grade, and Edmondson grade III and IV are high-grade, resulting in 21 low-grade and 25 high-grade HCCs for this study. Note that the clinical data has been used in the work of [4,6].

3.2   Performance of Local and Nonlocal Deep Feature

Table 1 showed the characterization performance of local, non-local and the proposed local and non-local fusion of deep features from the arterial phase of Contrast-enhanced MR in 2D and 3D, respectively. First, it can be found that 3D deep feature outperformed 2D deep feature either in local or non-local circumstances for malignancy characterization of HCC, which demonstrated that 3D CNN or 3D Non-local Neural network encoded sufficiently spatial information in volumetric data compared with 2D CNN or 2D Non-local Neural network. Furthermore, non-local deep feature showed better performance than local deep feature for malignancy characterization both in 2D and 3D, indicating that non-local deep feature may embed more image feature from vascularity and cellularity of neoplasm to characterize the aggressiveness of HCC. Finally, the proposed local and non-local deep feature fusion yielded best results both in 2D and 3D when taking advantage of local and non-local deep features.

3.3   Comparison of Deep Feature Fusion Methods

Table 2 showed the performance comparison of local and non-local deep feature fusion by direct concatenation, deep correlation model and the common and individual feature in 2D and 3D, respectively. Compared with the performance of local or non-local deep features in 2D and 3D as tabulated in Table 1, all the fusion methods could obtain improved results as shown in Table 2. Comparatively, the proposed fusion method based on common and individual feature analysis yielded better results than direct concatenation and deep correlation model both in 2D and 3D circumstances. Furthermore, the individual component between local and non-local deep features also yielded promising results

Table 1. Performance comparison of local, non-local and the proposed local and non-local fusion of deep features in 2D and 3D from the arterial phase of Contrast-enhanced MR(%).

	Accuracy	Sensitivity	Specificity
2D local	84.61 ± 7.25	87.60 ± 11.81	83.00 ± 17.19
2D non-local	86.32 ± 7.93	87.61 ± 12.16	87.24 ± 10.83
2D local and non-local fusion	89.74 ± 8.88	91.51 ± 8.55	88.17 ± 16.12
3D local	87.69 ± 7.05	86.96 ± 9.99	89.53 ± 11.58
3D non-local	90.00 ± 6.01	86.95 ± 9.35	93.42 ± 8.24
3D local and non-local fusion	93.16 ± 4.36	94.14 ± 8.92	91.30 ± 11.26

for malignancy characterization of HCC, especially in 3D. Specifically, the common feature yielded slightly better results than that of the deep correlation model, demonstrating that the common component recovered by the common and individual feature analysis has more advantage than that from the canonical correlation analysis, which is consistent with the previous finding in [13].

Table 2. Performance comparison of local and non-local deep feature fusion in 2D and 3D by direct concatenation, deep correlation model and the common and individual feature analysis(%).

	Accuracy	Sensitivity	Specificity
2D direct concatenation	88.03 ± 8.20	92.44 ± 11.77	86.01 ± 12.37
2D deep correlation model	88.89 ± 9.67	93.94 ± 6.92	85.81 ± 14.07
2D common feature	89.74 ± 6.28	92.35 ± 6.98	89.09 ± 10.16
2D individual feature	88.89 ± 8.20	92.70 ± 6.71	87.39 ± 13.13
2D common and individual feature analysis	90.60 ± 4.83	93.77 ± 7.35	87.50 ± 9.43
3D direct concatenation	91.45 ± 7.64	93.52 ± 9.80	90.22 ± 12.02
3D deep correlation model	90.60 ± 6.04	87.70 ± 9.40	91.45 ± 11.22
3D common feature	91.45 ± 5.67	91.16 ± 8.71	91.91 ± 7.64
3D individual feature	92.31 ± 6.28	91.78 ± 9.84	91.30 ± 12.71
3D common and individual feature analysis	93.16 ± 4.36	94.14 ± 8.92	91.30 ± 11.26

4   Conclusion

The proposed local and non-local deep feature fusion model yields superior performance for malignancy characterization of HCC in comparison of local deep feature, non-local deep feature, and the fusion methods of direct concatenation and deep correlation model, providing a novel strategy for the biological aggressiveness prediction and treatment planning of neoplastic diseases.

References

1. Park, J.W., Chen, M., Colombo, M., et al.: Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE study. Liver Int. **35**(9), 2155–2166 (2015)
2. Bruix, J., Sherman, M.: Management of hepatocellular carcinoma. Hepatology **42**, 1208–1236 (2005)
3. Nishie, A., Tajima, T., Asayama, Y., et al.: Diagnostic performance of apparent diffusion coefficient for predicting histological grade of hepatocellular carcinoma. Eur. J. Radiol. **80**(2), e29–e33 (2011)
4. Zhou, W., Zhang, L., Wang, K., et al.: Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced MR images. J. Magn. Reson. Imaging **45**(5), 1476–1484 (2017)
5. Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**(9), 60–88 (2017)
6. Wang, Q., Zhang, L., Xie, Y., Zheng, H., Zhou, W.: Malignancy characterization of hepatocellular carcinoma using hybrid texture and deep feature. In: Proceedings of the 24th IEEE International Conference Image Processing, pp. 4162–4166 (2017)
7. Wang, X., Girshick, R., Gupta A., He K.: Non-local neural networks. arXiv:1711.07971 (2017)
8. Setio, A.A.A., Ciompi, F., Litjens, G., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging **35**(5), 1160–1169 (2016)
9. Ciompi, F., de Hoop, B., Van Riel, S.J., et al.: Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the box. Med. Image Anal. **26**, 195–202 (2015)
10. Yao, J., Zhu, X., Zhu, F., Huang, J.: Deep correlational learning for survival prediction from multi-modality data. MICCAI 2017. LNCS, vol. 10434, pp. 406–414. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66185-8_46
11. Wang, A, Cai, J, Lu, J, Cham, T.: MMSS: multi-modal sharable and specific feature learning for RGB-D object recognition. In: IEEE International Conference on Computer Vision, pp. 1125–1133 (2015)
12. Wang, Z., Lin, R., Lu, J., Feng, J., Zhou, J.: Correlated and individual multi-modal deep learning for RGB-D object recognition. arXiv:1604.01655v2 [cs.CV] (2016)
13. Panagakis, Y., Nicolaou, M.A., Zafeiriou, S, Pantic, M.: Robust correlated and individual component analysis. IEEE Trans. Pattern Anal. Mach. Intel. **38**(8), 1665–1678 (2016)
14. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. arXiv:1412.6980 [cs.LG] (2014)