



DLNLF-net: Denoised local and non-local deep features fusion network for malignancy characterization of hepatocellular carcinoma

Haoyuan Huang^{a,1}, Yanyan Xie^{a,1}, Guangyi Wang^b, Lijuan Zhang^c, Wu Zhou^{a,*}

^a School of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

^b Department of Radiology, Guangdong Provincial People's Hospital, Guangzhou 510080, China

^c Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ARTICLE INFO

Article history:

Received 1 February 2022

Revised 17 October 2022

Accepted 23 October 2022

Keywords:

Hepatocellular carcinoma

Non-local operation

Feature denoising

Convolutional neural networks

ABSTRACT

Introduction: Hepatocellular carcinoma (HCC) is a primary liver cancer with high mortality rate. The degree of HCC malignancy is an important prognostic factor for predicting recurrence and survival after surgical resection or liver transplantation in clinical practice. Currently, deep features obtained from data-driven machine learning algorithms have demonstrated superior performance in characterising lesion features in medical imaging processing. However, previous convolutional neural network (CNN)-based studies on HCC lesion characterisation were based on traditional local deep features. The aim of this study was to propose a denoised local and non-local deep features fusion network (DLNLF-net) for grading HCC.

Methods: Gadolinium-diethylenetriaminepentaacetic-acid-enhanced magnetic resonance imaging data of 117 histopathologically proven HCCs were collected from 112 patients with resected HCC between October 2012 and October 2018. The proposed DLNLF-net primarily consists of three modules: feature denoising, non-local feature extraction, and bilinear kernel fusion. First, local feature maps were extracted from the original tumour images using convolution operations, followed by a feature denoising block to generate denoised local features. Simultaneously, a non-local feature extraction block was employed on the local feature maps to generate non-local features. Finally, the two generated features were fused using a bilinear kernel model to output the classification results. The dataset was divided into a training set (77 HCC images) and an independent test set (40 HCC images). Training and independent testing were repeated five times to reduce measurement errors. Accuracy, sensitivity, specificity, and area under the curve (AUC) values in the five repetitive tests were calculated to evaluate the performance of the proposed method.

Results: Denoised local features (AUC 89.19%) and non-local features (AUC 88.28%) showed better performance than local features (AUC 86.21%) and global average pooling features (AUC 87.1%) that were derived from a CNN for malignancy characterisation of HCC. Furthermore, the proposed DLNLF-net yielded superior performance (AUC 94.89%) than a typical 3D CNN (AUC 86.21%), bilinear CNN (AUC 90.46%), recently proposed local and global diffusion method (AUC 93.94%), and convolutional block attention module method (AUC 93.62%) for malignancy characterisation of HCC.

Conclusion: The non-local operation demonstrated a better capability of yielding global representation, and feature denoising based on the non-local operation achieved performance gains for lesion characterisation. The proposed DLNLF-net, which integrates denoised local and non-local deep features, evidently outperforms conventional CNN-based methods in the malignancy characterisation of HCC.

© 2022 Elsevier B.V. All rights reserved.

Abbreviations: HCC, Hepatocellular Carcinoma; MRI, Magnetic Resonance Imaging; CNN, Convolutional Neural Network; GAP, global average pooling; CCA, canonical correlation analysis; CBAM, convolutional block attention module; LGD, Local and Global Diffusion; FLOPs, floating-point operations; LSD, least significant difference.

* Corresponding author.

E-mail address: zhouwu@gzucm.edu.cn (W. Zhou).

¹ These authors contributed equally to this work.

1. Introduction

Primary liver cancer is the third leading cause of cancer-related death worldwide, with approximately 906,000 new cases and 830,000 deaths in 2020. Hepatocellular carcinoma (HCC) is the most common primary liver cancer (comprising 75–85% of cases) [1]. The degree of HCC malignancy is an important prognostic factor for predicting recurrence and survival after surgical resection or liver transplantation in clinical practice. HCC with a higher pathological grade indicates a higher recurrence rate after surgery and

shorter survival time of patients [2]. Clinically, HCC malignancy is usually determined by the pathological examination of small tumour samples obtained *via* biopsy before surgery. However, the small pieces of tissue obtained *via* a biopsy may not fully represent tumour heterogeneity. Moreover, biopsies may not always be performed safely [3]. Therefore, non-invasive imaging-based methods may be a promising alternative for detecting the overall malignancy of HCC.

Various imaging methods, including ultrasound (US) [4], magnetic resonance (MR) [5], and computed tomography (CT) [6], have been used to assess the histological grade of HCC by identifying reliable imaging features. Previous studies have shown that various imaging features are related to HCC malignancy, including vascularity, cell differentiation, and capillary perfusion [7]. Moreover, image texture analysis based on functions such as pixel correlation, grayscale, and spectrum is widely used for tumour characterisation [8]. In particular, the recent use of computer-aided methods in conjunction with texture analysis and machine learning-related technologies has shown significant potential in the assessment of malignant lesions [9].

Currently, deep features obtained from data-driven machine learning algorithms have demonstrated superior performance in characterising lesion features in medical image processing. The most successful deep learning model for data analysis is the convolutional neural network (CNN), which has shown superior performance in medical image analysis [10]. In terms of the liver, 2D CNNs have recently been used for grading [11] and classification of benign and malignant HCCs [12]. Three-dimensional CNNs have also recently been applied for liver tumour diagnosis [13]. We have also previously studied the performance of 2D and 3D CNNs based on contrast-enhanced MR imaging (MRI) [17,19] and diffusion-weighted MRI [14] in relation to HCC classification.

However, previous CNN-based studies on lesion characterisation of HCC were based on traditional local deep features. In such deep features, convolutional operations are building blocks based on a single local neighbourhood and global dependencies are often ignored [15]. We believe that establishing global dependencies on deep features is important for malignant characterisation of HCC. Recently, Vision Transformer (ViT) [18] has shown a more superior performance than the CNN-based model because of its powerful global dependency modelling capabilities, where the self-attention mechanism is the key to modelling global dependence. However, the large number of parameters and lack of inductive biases result in an inferior performance by the ViT when compared to that of CNN for small datasets [18]. Therefore, to effectively extract global feature representations on small datasets such as HCC, we introduced a non-local module based on the self-attention mechanism, which is a lightweight module with a higher local inductive bias [22].

However, MRI images often exhibit artefacts (folding, aliasing, motion ghosting) and noise due to hardware imperfections (e.g. magnetic field inhomogeneities), applied signal processing (e.g. image aliasing), or patient compliance (e.g. bulk or respiratory motion) [20]. A recent study revealed that minor disturbances in the image (e.g. noise or artefacts) may cause considerable noise in the deep features of the CNN system, which may result in false predictions [23]. Furthermore, a deep learning approach has recently been proposed to connect image denoising with high-level vision tasks [24]. This study indicates that denoising can remove noise while preserving important semantic details to produce correct high-level vision task results. Therefore, this study explored the effect of feature denoising on the performance of malignancy characterisation in HCC. To our knowledge, no previous study has considered denoising of deep features in convolutional networks to improve its characterisation performance.

In the present study, we propose a denoised local and non-local deep features fusion network (DLNLF-net) to improve the malignancy characterisation of HCC, in which the denoised local and non-local deep features are first extracted in parallel, followed by the application of a bilinear kernel model to fuse them for lesion characterisation. The contributions of this study are summarised as follows: (1) We introduce a lightweight non-local feature extraction module to extract global deep features, which can improve the malignancy characterisation performance of HCC when compared to using local deep features. (2) We propose a feature denoising module and are the first to demonstrate that the denoising of local deep features can improve the malignancy characterisation performance of HCC. (3) We propose a novel DLNLF-net for the malignancy characterisation of HCC, which integrates the denoised local and non-local deep features using a bilinear kernel model, thereby markedly outperforming conventional CNN-based methods.

2. Methods

2.1. Study population and clinical features

The local institutional review board approved this study, and informed consent was obtained from all patients. This retrospective study included 117 pathologically confirmed HCC lesions in 107 patients, identified between October 2012 and October 2018 (97 males and 10 females; age range: 27–71 years; mean age: 52.47 years). The inclusion criteria were as follows: (a) a new diagnosis of HCC followed by surgical resection was performed; (b) no prior treatment, including radiofrequency ablation, hepatectomy, percutaneous ethanol injection, or transcatheter arterial chemoembolisation, was conducted; (c) available routine clinical MR images were used; and (d) available histological reports of liver cancer were considered. HCC histological information was retrieved from the archived clinical histology report describing tumour presence, differentiation, size, microvascular invasion, surgical resection of the margin, and presence or absence of HCC cirrhosis or fibrosis. Of the 117 HCCs, 64 high-grade tumours corresponded to Edmondson grades III and IV, whereas 53 low-grade tumours corresponded to Edmondson grades I and II.

2.2. Magnetic resonance protocol

MRI was performed using a 3.0T MR scanner (SIGNA Excite HD 3.0T, GE Healthcare, Milwaukee, WI, USA) with an eight-channel phased-array body coil. The pre-contrast and three-phase post-contrast images were acquired using breath-hold axial liver acquisition with volume acceleration protocols with transverse T1-weighted dual-phase fast spoiled gradient-echo imaging (repetition time/time to echo, 4.25 ms/1.95 ms; flip angle, 90°; matrix, 512×512; slice thickness, 2.2 mm; interslice gap, 1 mm; field of view, 35–40 cm). Contrast-enhanced MR images were acquired after a rapid bolus injection of 0.025 mmol/kg body weight (0.1 ml/kg) gadolinium-diethylenetriaminepentaacetic acid (Magnevist, Bayer-Schering Pharma AG) at a rate of 2.5 ml/s, which was immediately followed by a 15 ml saline flush through a power injector at a rate of 2 ml/s. Arterial, portal venous, and delayed phase images were acquired at 20–25, 60, and 180 s after contrast agent injection, respectively. Fig. 1 shows representative contrast-enhanced MR images of a 59-year-old man with pathologically confirmed low-grade HCCs (grade II).

2.3. Proposed framework

Fig. 2 illustrates the proposed DLNLF-net for 2D and 3D lesion characterisation, adopting two CNN architectures based on a

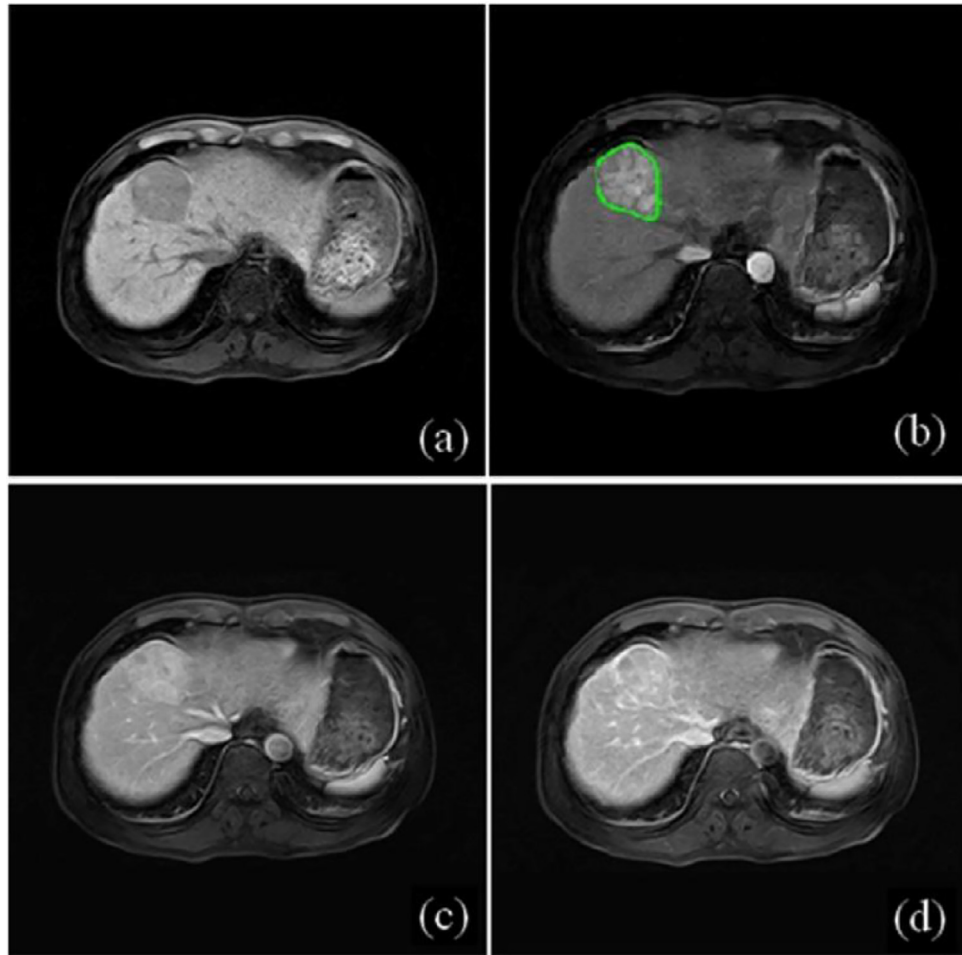


Fig. 1. Axial views of contrast-enhanced magnetic resonance images in a 59-year-old man with pathologically confirmed low-grade hepatocellular carcinoma: (a) Pre-contrast phase; (b) Arterial phase; (c) Portal venous phase; (d) Delayed phase. The tumour region is marked in green in (b).

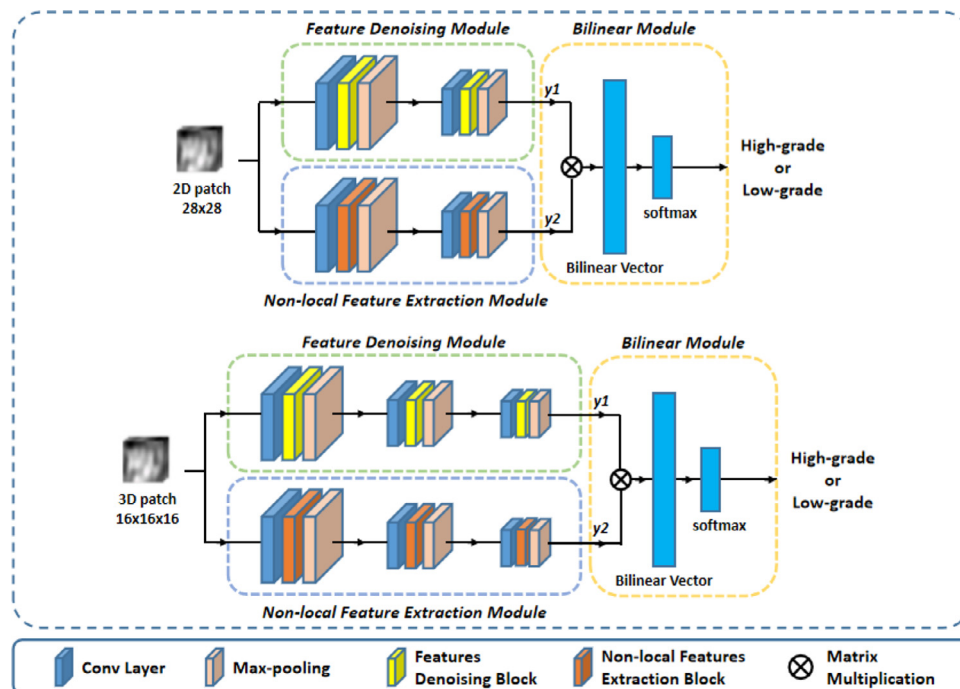


Fig. 2. Denoised local and non-local deep features fusion network (DLNLF-net) for 2D (upper panel) and 3D (lower panel) lesion characterisation.

simplified Visual Geometry Group architecture in parallel, to separately extract denoised and global deep features, and then combine them for lesion characterisation. For the 2D framework, the composition of the feature denoising module is as follows: two convolutional layers of size 3×3 each with a stride of 1 and 64 kernels, a feature denoising block, a 2×2 max-pooling layer, two convolutional layers of size 3×3 each with a stride of 1 and 128 kernels, a feature denoising block, and a 2×2 max-pooling layer. The composition of the non-local feature extraction module is essentially the same as that of the feature denoising module. For the 3D framework, the composition of the feature denoising module is as follows: two convolutional layers of size $3 \times 3 \times 3$ each with a stride of 1 and 32 kernels, a feature denoising block, a $2 \times 2 \times 2$ max-pooling layer, two convolutional layers of size $3 \times 3 \times 3$ each with a stride of 1 and 64 kernels, a feature denoising block, a $2 \times 2 \times 2$ max-pooling layer, two convolutional layers of size $3 \times 3 \times 3$ each with a stride of 1 and 128 kernels, a feature denoising block, and a $2 \times 2 \times 2$ max-pooling layer. The composition of the non-local feature extraction module is essentially the same as that of the feature denoising module. Finally, the denoised local and global feature outputs from the feature denoising and non-local feature extraction modules are fused using a bilinear kernel model to output the classification results with a softmax layer. The main novelty of the proposed DLNLF-net is that we simultaneously considered the denoised local features, global features, and their fusion to improve the performance of lesion characterisation. Here, we employed conventional cross-entropy as the loss function for the classification of low- and high-grade HCCs. The details of the three modules are introduced in the following sections.

2.4. Non-local mean operations for feature denoising and non-local feature extraction

Non-local mean operations were originally proposed for image denoising [25], which compute a denoised feature map y from an input feature map x by determining the weighted mean of features in all spatial locations L :

$$y_i = \frac{1}{\mathcal{C}(x)} \sum_{j \in L} f(x_i, x_j) x_j, \quad (1)$$

where, i is the index of the location to be computed and j is the index of all possible locations in the feature map x . The weighting function $f(x_i, x_j)$ calculates a scalar that represents the similarity between i and j . The response was normalised by the factor $\mathcal{C}(x) = \sum_{j \in L} f(x_i, x_j)$. The weighted average in Eq. (1) is greater than x_j . Therefore, denoising is performed directly on the input feature map x , and the correspondence between the feature channels in y and x is maintained.

In a non-local neural network, the non-local mean operation in deep learning networks is defined as [22]

$$y_i = \frac{1}{\mathcal{C}(x)} \sum_{j \in L} f(x_i, x_j) g(x_j), \quad (2)$$

where, $f(x_i, x_j)$ is the weighting function and $\mathcal{C}(x)$ is a normalisation factor, which is identical to that in Eq. (1). $g(x_j)$ is the unary function that computes the representation of the input feature map x at position j . Therefore, the weighted average in Eq. (2) is greater than the embedding of x_j , indicating non-local behaviour in Eq. (2) to generate a global deep feature y from the input feature map x .

To effectively implement non-local operations in the deep learning framework, the weighting functions $f(x_i, x_j)$ in Eqs. (1) and (2) are usually treated as an embedded Gaussian function [22]: $f(x_i, y_j) = e^{\theta(x_i)^T \phi(x_j)}$, where $\theta(x_i) = W_\theta x_i$ and $\phi(x_j) = W_\phi x_j$ are

two embeddings. For simplicity, the value of g in Eq. (2) is considered in the form of a linear embedding: $g(x_j) = W_g x_j$. W_θ , W_ϕ , and W_g are the three weight matrices to be learned, which are implemented as 1×1 and $1 \times 1 \times 1$ convolutions in the 2D and 3D modules, respectively. Fig. 3 shows the implementation of the feature denoising block and the non-local feature extraction block based on non-local mean operations in the 2D DLNLF-net. The components within the red rectangles in Fig. 3(a) and (b) illustrate the implementation of the non-local means that are measured as shown in Eqs. (1) and (2).

2.5. Bilinear kernel model

Bilinear CNNs have been proposed to combine two identical CNN features similar to a pooled outer product, which has been shown to outperform existing models on a variety of recognition tasks [26]. Inspired by the idea of generating a new representation using a pooled outer product of two designed features, we propose the fusion of the denoised local deep features and the non-local features of the tumour as a pooled outer product to achieve better characterisation performance. Recently, a bilinear kernel model was introduced to combine local and global features for video recognition, which exhibited high performance [15].

The feature fusion module used in the proposed network is illustrated on the right side of Fig. 2. This module combines the denoised local features y_1 from the denoised local feature block and the non-local features y_2 from the non-local feature extraction block. A bilinear vector B is obtained using the matrix output product $B = y_1 \cdot y_2$ followed by the process of normalisation, including a signed square root step $\text{sign}(B) \sqrt{|B|}$ and L_2 normalisation $B/\|B\|_2$ to improve the performance. The bilinear feature $B/\|B\|_2$ is a general-purpose image representation that can be passed through a linear and softmax layer to obtain class prediction.

2.6. Implementation

The proposed framework was implemented using the TensorFlow GPU platform, while training and testing were performed in a NVIDIA GTX1080 configuration. We implemented both the 2D and 3D modules of the proposed DLNLF network. Image resampling for data augmentation was applied to extract several patches as samples from each lesion. Please note that both the extracted 2D and 3D patches contained almost the entire tumour area. The details of the 2D and 3D data augmentations can be obtained from the supplementary material, which are similar to earlier reports [14,17,27]. The Adam algorithm was used to minimise the total loss function of the deep learning framework for each model. The learning rate was initially set to 1×10^{-5} , batch size was set to 64, and learning rate decay was 0.98. The source code of the proposed modules in this study is available (<https://github.com/Ksuriuri/DLNLF-net>).

3. Results

3.1. Dataset and evaluation metric

To demonstrate the performance of the proposed method for lesion characterisation, we performed comprehensive experiments on the malignancy characterisation of HCC using arterial-phase contrast-enhanced MRI. We selected HCC images in the arterial phase for analysis because the HCC ingested the contrast agent during the arterial phase, which enabled a high degree of recognition of malignant cells and blood vessels [28]. The experiment was repeated five times, with 40 HCCs randomly selected from 117 HCCs as the test set in each cycle; the remaining HCCs (77 HCCs) were used as the training set. Note that we ensured that the same

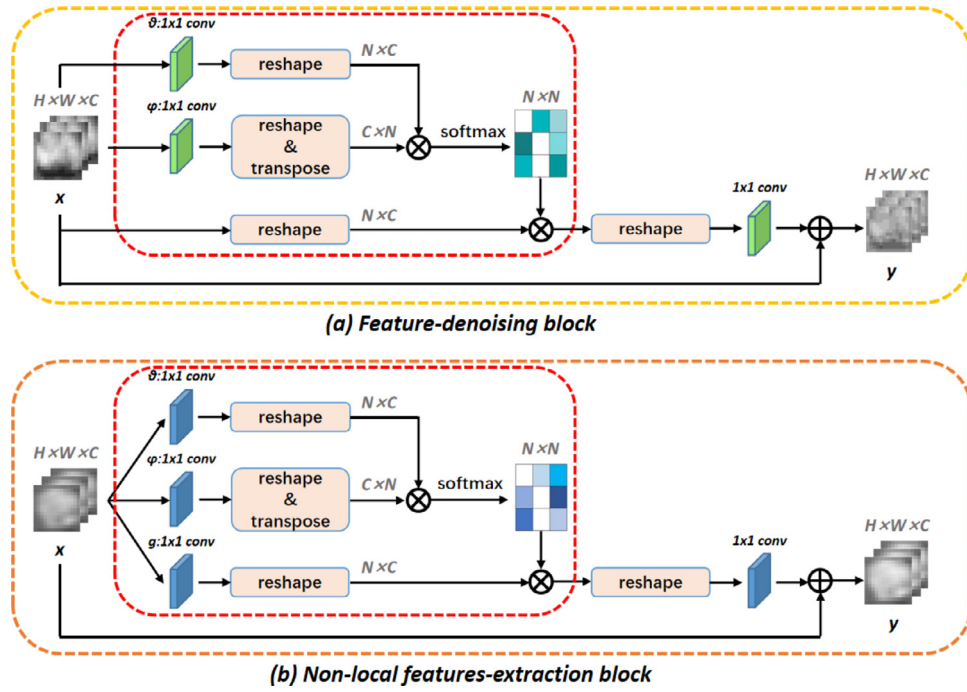


Fig. 3. (a) Feature denoising block and (b) non-local feature extraction block in the 2D DLNLF-net. Suppose the input tensor of the original image is $H \times W \times C$, where H and W denote the height and width of the input image, respectively, and C denotes the channel; then, $N = H \times W$. In our experiment, the input tensor of the 2D DLNLF-net is $28 \times 28 \times 32$. Note that the implementation of the two blocks in the 2D and 3D DLNLF-nets is straightforward, and the input tensor is $16 \times 16 \times 16 \times 64$ for the 3D DLNLF-net. The softmax function is implemented along dimension j to compute $\frac{1}{\sum_k} f(x_i, x_j)$.

training-test set partition was used in the different methods in each experiment. We performed a test on the test set for every 10 iterations and saved the results. The values of the metrics in the five replicated tests, including accuracy, sensitivity, specificity, and F1-score, were calculated and presented in the form of mean \pm standard deviation. The performance of the prediction model was evaluated using the following statistical metrics.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Here, TP, FP, TN, and FN represent the true-positive, false-positive, true-negative, and false-negative results, respectively.

The receiver operating characteristic (ROC) curve and area under the curve (AUC) values were also used to evaluate the characterisation performance. The statistical differences in the output probability values of the deep learning models between high- and low-grade HCCs were evaluated using independent Student's t-tests. Differences were considered to be statistically significant at P values less than 0.05.

We first implemented the baseline architecture of the CNN with conventional local convolutions for the task of malignancy differentiation of HCCs, and then extended it to the denoised local features and global non-local deep features via non-local operations. Specifically, we separately assessed the performance of the local features, conventional global average pooling (GAP) features [29], proposed non-local deep features, and denoised local deep features for lesion characterisation. Then, we compared the

performance of different fusion models to combine the denoised local and non-local deep features, including concatenation [17], canonical correlation analysis (CCA) [32], correlated and individual analysis (C+I) [33,34], and the adopted bilinear kernel model [28]. Furthermore, we compared the performance of the proposed DLNLF-net with that of state-of-the-art methods, including typical CNNs [13,16,17,19,21], local/global diffusion (LGD) [15], bilinear CNNs [26], convolutional block attention module (CBAM) [30], and local/non-local fusion C+I [31]. Note that both 2D and 3D CNNs based on the reported methods and the proposed DLNLF-nets were implemented for performance comparison. Subsequently, we elaborately compared the testing processes of different deep learning models for HCC classification. Finally, we presented the computational cost of different modules and adopted the least significant difference (LSD) method [35] to conduct multiple comparisons to ensure the overall level of type I error at a specific level.

To further verify the generalisation performance of our method, we also performed similar experiments on the diffusion-weighted imaging and brain tumour image segmentation datasets and obtained results consistent with the experiments in this paper, which can be obtained from the supplementary material.

3.2. Experiments on different modules: local, non-local, and global average pooling

Table 1 compares the performance of different modules when characterising the malignancy of HCC, both in 2D and 3D networks. When compared to the local features generated by the CNN baseline method, the GAP CNN with global features improved the accuracy from 81.02 to 82.74% in 2D characterisation and from 84.52 to 85.09% in 3D characterisation. Furthermore, the non-local CNN outperformed the GAP CNN, which obtained further improved performance with an accuracy of 83.98 and 86.32% in 2D and 3D characterisations, respectively. Finally, the denoised CNN obtained the largest performance gains, with an accuracy of 84.78% in 2D

Table 1

Characterisation performance comparison of different 2D and 3D convolutional neural network modules with the local, non-local, global average pooling, and denoised local components using a fixed test set (%). Δ represents the mean accuracy differences between the baseline and other methods.

	Model	Accuracy	Δ	Sensitivity	Specificity	F1-score	AUC (95% CI)	P
2D	CNN baseline [17]	81.02 \pm 3.65	–	80.20 \pm 6.82	79.95 \pm 7.19	79.36 \pm 6.51	83.36 \pm 3.23 (77.0, 89.7)	0.04
	GAP CNN [29]	82.74 \pm 3.04	1.72	82.35 \pm 5.75	82.90 \pm 7.67	81.84 \pm 5.54	84.2 \pm 3.01 (78.3, 90.1)	0.03
	Non-local CNN [22]	83.98 \pm 2.28	2.96	83.64 \pm 5.54	81.23 \pm 6.36	81.85 \pm 5.30	85.82 \pm 3.38 (79.2, 92.4)	0.02
	Denoised CNN [23]	84.78 \pm 2.12	3.76	84.23 \pm 4.89	82.79 \pm 5.12	82.88 \pm 4.66	86.32 \pm 2.83 (80.8, 91.9)	0.02
3D	CNN baseline [18]	84.52 \pm 3.21	–	82.38 \pm 5.74	84.84 \pm 5.21	82.74 \pm 5.45	86.21 \pm 3.61 (79.1, 93.3)	0.04
	GAP CNN [29]	85.09 \pm 3.10	0.57	86.46 \pm 5.01	84.92 \pm 4.33	85.13 \pm 4.75	87.10 \pm 3.21 (80.8, 93.4)	0.03
	Non-local CNN [22]	86.32 \pm 2.76	1.80	85.64 \pm 4.38	86.9 \pm 4.45	85.59 \pm 4.17	88.28 \pm 2.78 (82.8, 93.7)	0.02
	Denoised CNN [23]	87.27 \pm 2.44	3.18	88.92 \pm 4.29	86.87 \pm 3.21	87.42 \pm 4.06	89.14 \pm 2.52 (84.2, 94.1)	0.02

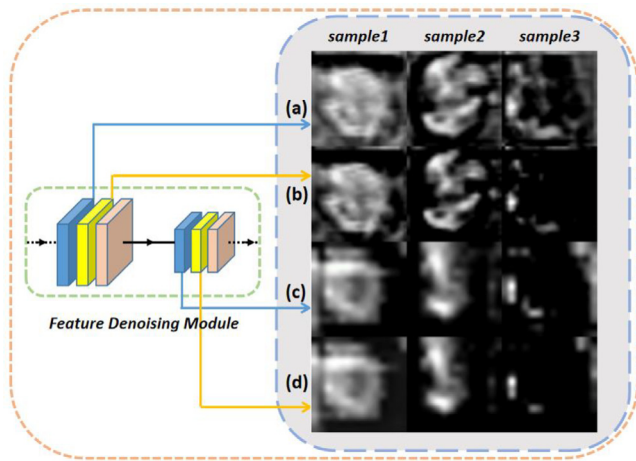


Fig. 4. Visualisation of local and denoised local features that were extracted from three different hepatocellular samples: (a) Local feature map from the first convolutional layer; (b) Denoised local feature map from the first convolutional layer; (c) Local feature map from the second convolutional layer; (d) Denoised local feature map from the second convolutional layer.

characterisation and 87.27% in 3D characterisation. Notably, the improvements were consistent in both the 2D and 3D modules, while the 3D modules always yielded better performance than the 2D modules. Furthermore, Fig. 4 shows the visualisation of the local and denoised local features that were extracted from three different HCCs. The local feature maps became smooth, and salient regions were highlighted by non-local operations.

3.3. Experiments based on different fusion methods: concatenation, canonical correlation analysis, correlated and individual analysis (C+I), and bilinear model

To demonstrate the superior fusion performance of the adopted bilinear model, Table 2 shows the 2D and 3D characterisation performance of different fusion methods when combining the denoised local and non-local deep features. The adopted bilinear model method resulted in a markedly higher performance for lesion characterisation than the previously proposed fusion methods. This clearly shows that the bilinear model outperforms the concatenation method by 4.49 and 5.84%, CCA by 0.67 and 2.53%, and C+I by 1.36 and 3.36% in terms of accuracy in 2D and 3D characterisations, respectively.

3.4. Comparison with the state-of-the-art methods: 3D convolutional neural network (CNN), convolutional block attention module, local/global diffusion, bilinear CNNs, and local/non-local C+I

Table 3 compares the performance of the different methods in the 2D and 3D networks. The proposed DLNLF-net showed notable improvements of 10.45 and 9.60% in accuracy as compared to the CNN baseline in the 2D and 3D modules, respectively. Furthermore, the proposed DLNLF-net also outperformed the current state-of-the-art methods, including CBAM, LGD, local/non-local C+I, and bilinear CNN, for both 2D and 3D characterisations. Specifically, the proposed DLNLF-net outperformed the bilinear CNN by 5.28 and 5.07%, CBAM by 2.15 and 2.98%, LGD by 2.71 and 1.79%, and local/non-local C+I by 3.13 and 2.58% in terms of accuracy in the 2D and 3D modules, respectively. Then, by applying the trained DLNLF-net models to the corresponding test set. The confusion matrix for our model in terms of 2D and 3D characterisations was determined, as shown in Table 4.

3.5. Detailed study of the testing process

Fig. 5 shows the accuracy, loss, and corresponding ROC curves of several typical models for the characterisation of HCC malignancy. The proposed DLNLF-net consistently exhibited a superior performance than the CNN baseline method and the models that extracted non-local features with CNN, denoised local features with CNN, and local and non-local deep features that are fused by the C+I method in the testing procedure. Fig. 6 shows two representative cases in which the proposed method made accurate predictions, whereas most of the other methods failed.

3.6. Computational cost

It should be noted that adding the non-local/denoising module to the CNN baseline module does not significantly increase the number of parameters and calculations. Table 5 shows a comparison of floating-point operations (FLOPs), trainable parameters, and inference time for different 3D models. When compared to the CNN baseline model, the increase in parameters and FLOPs after adding the non-local or denoising module is significantly small; however, the inference time is greater than that of the CNN baseline. This is because the non-local/denoising modules include many reshaping operations, which cause a disproportionate increase in time per iteration corresponding to the increase in FLOPs. In com-

Table 2

Characterisation performance comparison of different fusion methods that combine the denoised local and non-local deep features in the 2D and 3D networks using a fixed test set (%). Δ represents the mean accuracy differences between the baseline and the other methods.

	Method	Accuracy	Δ	Sensitivity	Specificity	F1-score	AUC (95% CI)	P
2D	Concatenation [17]	86.98 \pm 3.78	–	84.90 \pm 5.54	87.60 \pm 4.21	85.50 \pm 5.24	87.13 \pm 3.34 (80.6, 93.7)	0.02
	C+I [33,34]	88.34 \pm 2.48	1.36	88.82 \pm 3.75	87.55 \pm 4.00	87.69 \pm 3.57	89.53 \pm 3.01 (83.6, 95.4)	0.00
	CCA [32]	87.65 \pm 3.97	0.67	86.46 \pm 4.23	88.21 \pm 5.20	86.68 \pm 4.05	88.80 \pm 3.23 (82.5, 95.1)	0.00
	Bilinear (adopted) [28]	91.47 \pm 2.86	4.49	91.87 \pm 3.68	90.67 \pm 3.94	90.88 \pm 3.51	92.35 \pm 2.08 (88.3, 96.4)	0.00
3D	Concatenation [17]	88.28 \pm 3.91	–	86.66 \pm 4.82	89.57 \pm 4.73	87.45 \pm 4.59	89.11 \pm 3.49 (82.3, 96.0)	0.02
	C+I [33,34]	91.54 \pm 3.31	3.26	92.34 \pm 3.95	90.86 \pm 3.13	91.23 \pm 3.74	92.02 \pm 3.55 (85.1, 98.9)	0.00
	CCA [32]	90.81 \pm 3.22	2.53	90.94 \pm 3.20	89.82 \pm 3.47	89.95 \pm 3.05	91.46 \pm 3.00 (85.6, 97.3)	0.00
	Bilinear (adopted) [28]	94.12 \pm 2.02	5.84	94.53 \pm 2.97	93.28 \pm 2.76	93.61 \pm 2.82	94.89 \pm 1.98 (91.0, 98.8)	0.00

Table 3

Characterisation performance comparison of different methods in 2D and 3D networks using the fixed test set (%). Δ represents the mean accuracy differences between the baseline and other methods.

	Method	Accuracy	Δ	Sensitivity	Specificity	F1-score	AUC (95% CI)	P
2D	2D CNN baseline [15–19]	81.02 \pm 3.65	–	80.20 \pm 6.82	79.9 \pm 7.19	79.26 \pm 6.51	83.36 \pm 3.23 (77.0, 89.7)	0.04
	CBAM + CNN [30]	89.32 \pm 3.30	8.30	87.43 \pm 3.09	89.23 \pm 4.23	87.72 \pm 2.96	91.68 \pm 2.10 (87.6, 95.8)	0.00
	Local/Non-local C+I [31]	88.34 \pm 2.48	7.32	88.82 \pm 3.75	87.55 \pm 4.00	87.69 \pm 3.57	89.53 \pm 3.01 (83.6, 95.4)	0.00
	Local and Global Diffusion (LGD) [21]	88.76 \pm 3.25	7.74	89.37 \pm 3.23	88.56 \pm 3.70	88.48 \pm 3.08	90.77 \pm 2.34 (86.2, 95.4)	0.00
	Bilinear CNNs [26]	86.19 \pm 2.67	5.17	87.93 \pm 3.13	85.43 \pm 3.26	86.19 \pm 2.98	87.23 \pm 2.74 (81.9, 92.6)	0.00
	DLNLF-nets (proposed)	91.47 \pm 2.86	10.45	91.87 \pm 3.68	90.67 \pm 3.94	90.88 \pm 3.51	92.35 \pm 2.08 (88.3, 96.4)	0.00
3D	3D CNN baseline [18]	84.52 \pm 3.21	–	82.38 \pm 5.74	84.84 \pm 5.21	82.74 \pm 5.45	86.21 \pm 3.61 (79.1, 93.3)	0.04
	CBAM + CNN [30]	91.14 \pm 3.35	6.62	92.52 \pm 3.72	90.31 \pm 3.69	91.05 \pm 3.54	93.62 \pm 2.15 (89.4, 97.8)	0.01
	Local/Non-local C+I [31]	91.54 \pm 3.31	7.02	92.34 \pm 3.95	90.86 \pm 3.13	91.23 \pm 3.74	92.02 \pm 3.55 (85.1, 98.9)	0.00
	Local and Global Diffusion (LGD) [21]	92.33 \pm 3.78	7.81	91.29 \pm 3.42	93.00 \pm 3.67	91.74 \pm 3.26	93.94 \pm 2.74 (88.6, 99.3)	0.00
	Bilinear CNNs [26]	89.05 \pm 2.57	4.53	89.29 \pm 3.83	87.73 \pm 3.33	88.03 \pm 3.64	90.46 \pm 2.31 (86.0, 95.0)	0.00
	DLNLF-nets (proposed)	94.12 \pm 2.02	9.6	94.53 \pm 2.97	93.28 \pm 2.76	93.61 \pm 2.82	94.89 \pm 1.98 (91.0, 98.8)	0.00

Table 4

Confusion matrix of the denoised local and non-local deep features fusion network model for 2D and 3D characterisations.

Predicted class	2D DLNLF-net	Actual class		3D DLNLF-nets	Actual class	
		Low-grade	High-grade		Low-grade	High-grade
	Low-grade	19	2	Low-grade	20	1
	High-grade	2	17	High-grade	1	18

Table 5

Comparison of floating-point operations, trainable parameters, and time per iteration for different 3D models.

	FLOPs	Trainable parameters	Inference time (ms)
CNN baseline	1.08 \times 10 ⁶	0.54 \times 10 ⁶	38.00 \pm 6.02
Non-local CNN	1.17 \times 10 ⁶	0.58 \times 10 ⁶	81.79 \pm 7.11
Denoised CNN	1.17 \times 10 ⁶	0.58 \times 10 ⁶	81.11 \pm 6.68
DLNLF-net (proposed)	1.35 \times 10 ⁶	0.67 \times 10 ⁶	119.64 \pm 3.72

parison, the proposed DLNLF-net has relatively larger FLOPs, trainable parameters, and inference time than the CNN baseline owing to the use of two CNN architectures and non-local/denoising modules.

3.7. Multiple comparisons

Since there are many pairwise comparisons of methods, to ensure the overall level of the type I error at a specific level, we adopted the least significant difference (LSD) method [35] to conduct multiple comparisons. LSD method is used in ANOVA to create confidence intervals for all pairwise differences between factor level means while controlling the individual error rate to a specific significance level. Fig. 7 shows the results of multiple comparisons between different models using the LSD method, which correspond to the results listed in Tables 1–3. Most pairwise comparisons showed a statistically significant difference, indicating that the two models are indeed significantly different, and the probability of making a type I error in this judgement is only 5%.

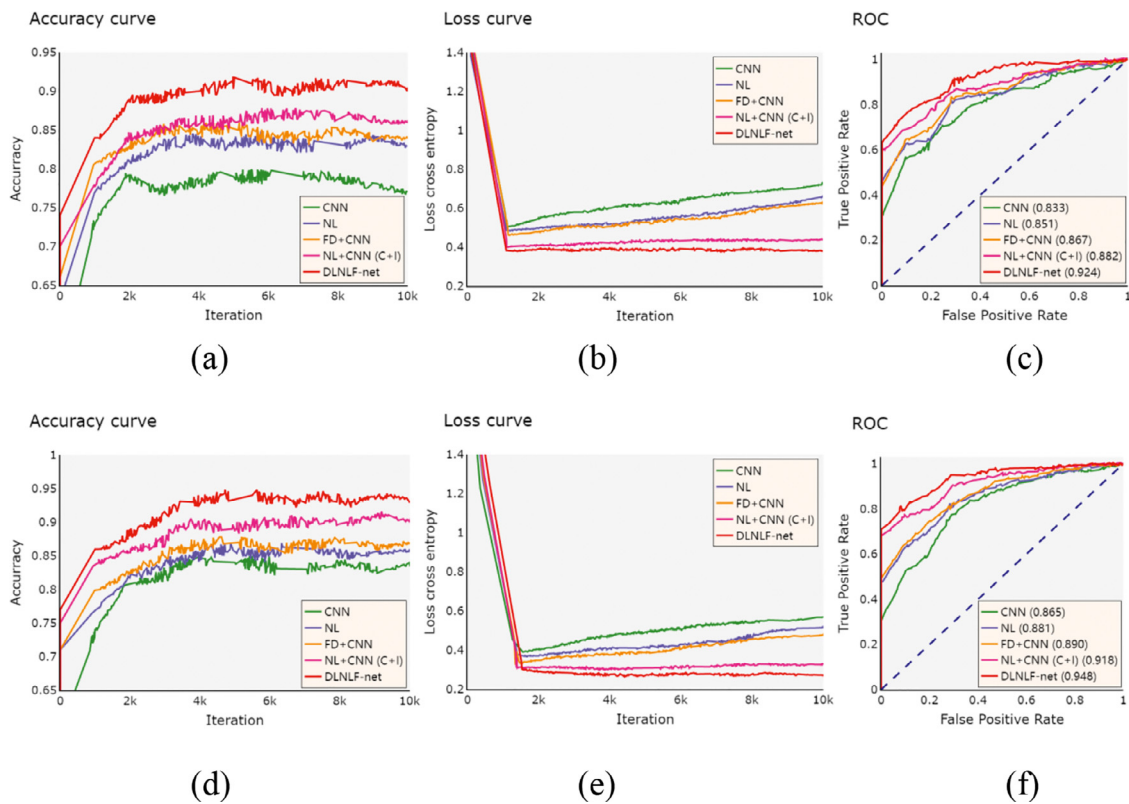


Fig. 5. Accuracy, loss, and receiver operating characteristic curves for the different methods in the (a–c) 2D and (d–f) 3D modules, respectively. The convolutional neural network (CNN) is the baseline model, NL represents the non-local features with CNN, FD+CNN represents the denoised local features with CNN, NL+CNN (C+I) represents the local and non-local deep features fused by the correlated and individual feature method [31], and DLNLF-net is the proposed method.

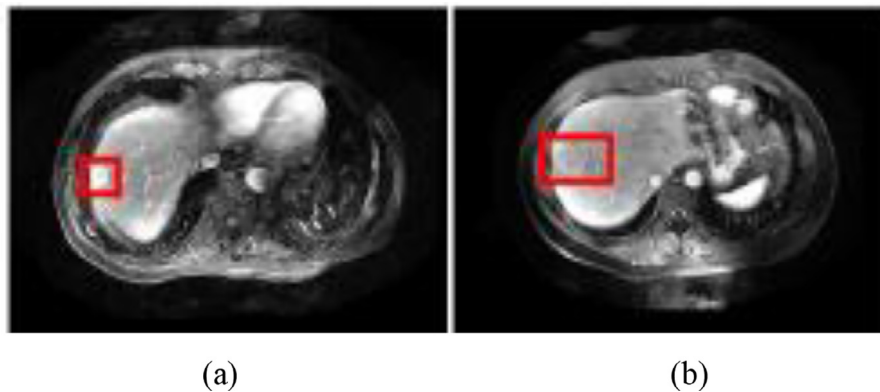


Fig. 6. Two representative cases. (a) A 72-year-old man with pathologically confirmed hepatocellular carcinoma (HCC; Grade II; 17 mm diameter). The proposed DLNLF-net and convolutional block attention module successfully identified the low-grade tumour, while the other four comparison methods made false predictions. (b) A 27-year-old male with a low HCC (Grade II; 40 mm diameter). The proposed DLNLF-net accurately identified the tumour, whereas all the other methods made false predictions.

In contrast, several comparisons showed no statistically significant difference, indicating that the apparent difference may have been caused by chance. Overall, the proposed DLNLF-net and the previously proposed methods showed statistically significant differences.

3.8. Ablation study on residual network

To demonstrate the effectiveness of the proposed modules in different backbone networks, we experimentally assessed the performance of an 18-layer residual network (ResNet) [37]. As listed in Table 6, the non-local and denoised modules can improve the characterisation performance when compared to the ResNet baseline,

whereas the DLNLF-net-based ResNet yields an even better performance.

4. Discussion

This study demonstrated that global features outperform local deep features in lesion characterisation (Table 1 and Figs. 5 and 7). As global deep features reflect the global dependency of features within the lesion, they are considered to be more representative and robust to large variations and complexities in image content. Recent computer vision studies have used GAP to yield global features [15,29,30]. However, our results indicate that the global features obtained by the non-local operation can achieve even better performance than GAP for lesion characterisation, implying that

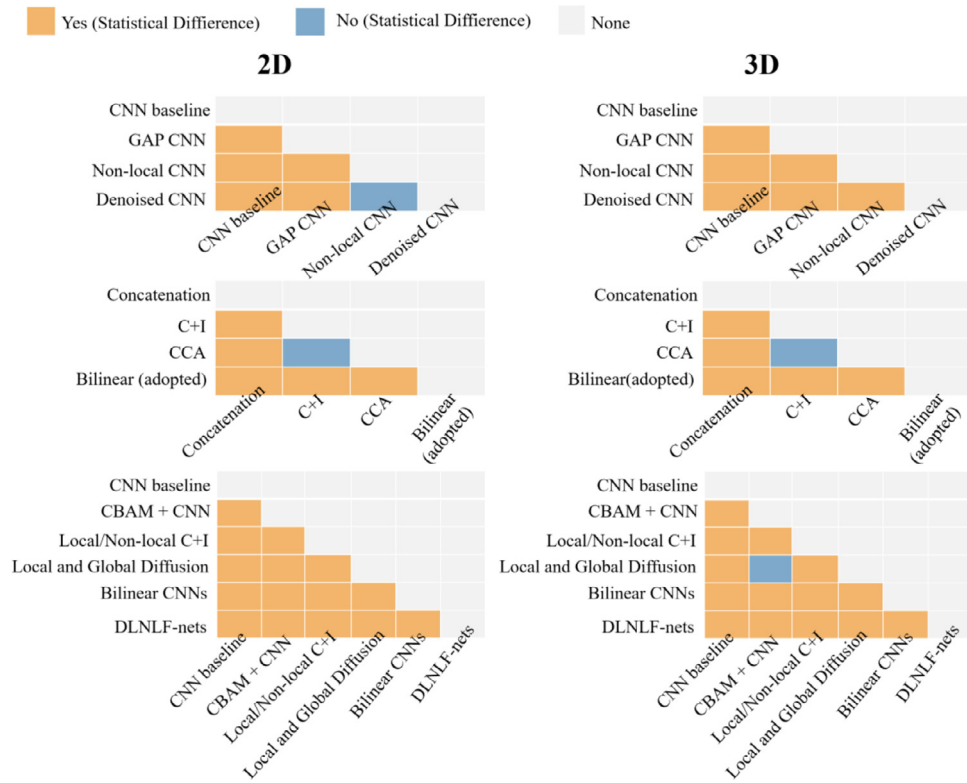


Fig. 7. Multiple comparisons between different methods corresponding to the results listed in Tables 1–3. Each small square indicates that the corresponding model was compared in pairs to evaluate whether there are statistical differences between them. Orange indicates that there is a statistical difference, blue indicates that there is no statistical difference, and grey represents that the results are not applicable.

Table 6

Comparison of 2D characterisation using different models based on a residual network baseline. Δ represents the mean accuracy differences between the baseline and the other methods.

Method	Accuracy	Δ	Sensitivity	Specificity	F1-score	AUC (95% CI)	P
ResNet baseline	81.46 ± 3.37	–	81.29 ± 5.24	79.17 ± 4.54	79.57 ± 4.97	83.50 ± 3.94 (75.8, 91.2)	0.03
Non-local+ResNet	84.47 ± 2.85	3.01	83.53 ± 4.77	82.07 ± 4.18	82.15 ± 4.53	85.60 ± 2.99 (79.7, 91.5)	0.02
Denoised +ResNet	83.89 ± 2.11	2.43	83.31 ± 3.37	81.97 ± 6.52	81.98 ± 3.26	86.10 ± 3.13 (80.0, 92.2)	0.02
DLNLF-netResNet	91.14 ± 2.12	9.68	92.11 ± 3.52	90.38 ± 4.11	90.86 ± 3.36	92.30 ± 2.87 (86.7, 97.9)	0.00

non-local operations have a better capability to provide a global representation for lesion characterisation. We believe that this improvement may be partially ascribed to the caption of long-range dependencies of non-local modules by computing interactions between any two positions, which is reported to be a self-attention mechanism to concentrate on dominant features [22,36].

In the present study, we further demonstrated that the denoised local deep features outperformed the local deep features for lesion characterisation (Table 1 and Figs. 5 and 7). Local features are susceptible to noise and image artefacts, which may lead to erroneous predictions by deep learning systems [23]. The adoption of a feature denoising block for lesion characterisation was inspired by a recently proposed method for improving adversarial robustness by feature denoising [23]. To the best of our knowledge, no previous study has demonstrated the significance of local feature denoising in deep learning for lesion characterisation. More importantly, the visualisation of feature maps in the model (Fig. 4) indicated that the denoised local feature maps became smooth, and salient regions were highlighted by non-local operations.

Furthermore, the bilinear model, which is a fusion method, was compared with conventional fusion methods, such as concatena-

tion, CCA, and the recently proposed C+I method (Table 2). Intuitively, the simple concatenation process has not explored the interactive relationship between the two vectors. CCA was proposed to extract the correlated components between the two vectors for characterisation. However, these vectors may be highly nonlinear. The recently proposed C+I method is intended to extract both the correlated and individual components between the two vectors for characterisation, whereas the correlated and individual vectors may not be optimally combined for lesion characterisation. In comparison, the adopted bilinear model fusion is based on the idea of combining two features using the pooled outer product, which has been shown to be the best for lesion characterisation because of the high efficacy of outer product representations.

The proposed method outperformed the existing methods for lesion characterisation, which can be attributed to the superiority of simultaneously considering denoised local features, global features, and their optimal fusion for lesion characterisation (Table 3). In contrast, LGD uses GAP to extract global information, which is inferior to non-local features (Table 1). CBAM also uses an attention mechanism to focus on the important features for characterisation. Channel attention and spatial attention are two sequential

submodules in the CBAM. In contrast, the proposed denoised local and non-local deep features are extracted in parallel and subsequently fused using the bilinear model method, which results in better performance. Finally, the proposed DLNLF-net also outperformed the original bilinear CNN [28], indicating that the combination of denoised local and non-local deep features is superior to the combination of two identical local deep features using the same outer product for lesion characterisation. This comparison shows that the improvement was due to the integration of non-local blocks.

Although the proposed method achieved marked performance improvements in tumour characterisation, there are still some limitations that should be addressed in the future. First, this study uses a small data set from a single site, which may raise concerns about the robustness of the results. In future, it is worthwhile to integrate multi-site data with harmonized magnetic resonance protocols to expand the dataset and conduct more extensive comparative experiments with the state-of-the-art methods [38,39]. Second, non-local neural networks are computationally expensive (Table 5) and lack the mechanism of encoding rich structural information [40]. Therefore, it is desirable to explore an efficient and robust implementation of non-local blocks for lesion characterisation. Moreover, the proposed method was effective for feature learning in a single modality, whereas multiple modalities are widely adopted to generate complementary improvements when compared to a single modality [41]. Future work should focus on extending the proposed method to include multimodal feature learning.

5. Conclusion

This study indicated that global deep features and denoised local deep features obtained better performance gains for lesion characterisation as compared to the local deep features that are used by existing methods. Moreover, the fusion of global and denoised local deep features with the bilinear model method could further improve the performance of lesion characterisation, outperforming the 3D CNN baseline and other recently proposed methods. The proposed module can be generalised and integrated into a CNN-based architecture to improve the performance of lesion characterisation in clinical practice.

Further research is necessary to explore a non-local module that can encode richer structural information or has less computational overhead. Furthermore, extending the proposed method to multimodal feature learning is a promising future work.

Grant Support

This research is partially sponsored by the grants from National Natural Science Foundation of China (81771920).

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgements

The authors highly thank the Department of Radiology, Guangdong General Hospital for providing MR images and clinical histology reports of HCCs for this research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2022.107201.

References

- [1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin.* (2021) Feb 4 (2021), doi:10.3322/caac.21660.
- [2] J. Bruix, M. Sherman, P.G. Committee, et al., American association for the study of liver diseases management of hepatocellular carcinoma: an update, *Hepatology* 53 (2005) 1020–1022.
- [3] B. Terjung, I. Lemnitzer, F.L. Dumoulin, et al., Bleeding complications after percutaneous liver biopsy, *Digestion* 67 (3) (2003) 138–145.
- [4] H.J. Jang, T.K. Kim, P.N. Burns, S.R. Wilson, Enhancement patterns of hepatocellular carcinoma at contrast-enhanced us: comparison with histologic differentiation, *Radiology* 244 (3) (2007) 898–906.
- [5] C.D. Witjes, F.E. Willemssen, J. Verheij, et al., Histological differentiation grade and microvascular invasion of hepatocellular carcinoma predicted by dynamic contrast-enhanced mri, *J. Magn. Reson. Imaging* 36 (3) (2012) 641–647.
- [6] H.Y. Kim, J.Y. Choi, C.W. Kim, et al., Gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid-enhanced magnetic resonance imaging predicts the histological grade of hepatocellular carcinoma only in patients with child-pugh class a cirrhosis, *Liver Transpl.* 18 (7) (2012) 850–857.
- [7] A. Kitao, Y. Zen, O. Matsui, et al., Hepatocellular carcinoma: signal intensity at gadoxetic acid-enhanced mr imaging—correlation with molecular transporters and histopathologic features, *Radiology* 256 (3) (2010) 817–826.
- [8] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, *Radiology* 278 (2) (2015) 563–577.
- [9] H.P. Chan, D. Wei, M.A. Helvie, et al., Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space, *Phys. Med. Biol.* 40 (5) (1995) 857–876.
- [10] D.R. Sarvamangala, R.V. Kulkarni, Convolutional neural networks in medical image understanding: a survey, *Evol. Intell.* 15 (2021) 1–22 3 January.
- [11] M. Chen, B. Zhang, W. Topatana, et al., Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning, *NPJ Precis. Oncol.* 4 (1) (2020) 1–7.
- [12] P.M. Oestmann, C.J. Wang, L.J. Savic, et al., Deep learning-assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) versus non-HCC on contrast-enhanced MRI of the liver, *Eur. Radiol.* 31 (7) (2021) 4981–4990.
- [13] C.A. Hamm, C.J. Wang, L.J. Savic, et al., Deep learning for liver tumor diagnosis part I: development of a convolutional neural network classifier for multi-phasic MRI, *Eur. Radiol.* 29 (2019) 3338–3347.
- [14] W. Zhou, G. Wang, G. Xie, L. Zhang, Grading of hepatocellular carcinoma based on diffusion weighted images with multiple b-values using convolutional neural networks, *Med. Phys.* 46 (9) (2019) 3951–3960.
- [15] Z. Qiu, T. Yao, C.W. Ngo, et al., Learning Spatio-Temporal Representation With Local and Global Diffusion, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE*, 2019, pp. 12056–12065.
- [16] K. Yasaka, H. Akai, O. Abe, S. Kiryu, Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study, *Radiology* 286 (3) (2018) 887–896.
- [17] T. Dou, L. Zhang, W. Zhou, 3D deep feature fusion in contrast-enhanced MR for malignancy characterization of hepatocellular carcinoma, in: *Proceedings of the 15th International Symposium on Biomedical Imaging*, 2018, pp. 29–33.
- [18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, ... N. Houlsby, in: *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*, International Conference on Learning Representations, 2020, September.
- [19] Q. Wang, L. Zhang, Y. Xie, H. Zheng, W. Zhou, Malignancy characterization of hepatocellular carcinoma using hybrid texture and deep features, in: *Proceedings of the International Conference on Image Processing*, 2017, pp. 4162–4166.
- [20] T. Küstner, S. Gatidis, A. Liebgott, M. Schwartz, L. Mauch, P. Martirosian, H. Schmidt, N.F. Schwenzer, K. Nikolaou, F. Bamberg, B. Yang, F. Schick, A machine-learning framework for automatic reference-free quality assessment in MRI, *Magn. Reson. Imaging* 53 (2018) 134–147 November.
- [21] K. Yasaka, H. Akai, A. Kunitatsu, O. Abe, S. Kiryu, Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images, *Radiology* 287 (1) (2018) 146–155.
- [22] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.
- [23] C. Xie, Y. Wu, L.V.D. Maaten, A. Yuille, K. He, Feature denoising for improving adversarial robustness, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2019, pp. 501–509.
- [24] D. Liu, B. We, J. Jiao, et al., Connecting image denoising and high-level vision tasks via deep learning, *IEEE Trans. Image Process.* 15 (1) (2020) 3695–3706.
- [25] A. Buades, B. Coll, J. Morel, A non-local algorithm for image denoising, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2, 2005, pp. 60–65.
- [26] T.Y. Lin, A. RoyChowdhury, S. Maji, Bilinear cnn models for fine-grained visual recognition, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2015, pp. 1449–1457.

- [27] H.R. Roth, L. Lu, J. Liu, et al., Improving computer-aided detection using convolutional neural networks and random view aggregation, *IEEE Trans. Med. Imag* 35 (5) (2015) 1170–1181.
- [28] W. Zhou, L. Zhang, K. Wang, et al., Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced mr images, *J. Magn. Reson. Imaging* 45 (5) (2017) 1476–1484.
- [29] M. Lin, Q. Chen, S. Yan, Network in network, in: *Proceedings of the 2nd International Conference of Learning Representations*, 2014 April.
- [30] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: convolutional block attention module, in: *Proceedings of the European Conference on Computer Vision*, 2018, pp. 3–19.
- [31] T. Dou, L. Zhang, H. Zheng, W. Zhou, Local and non-local deep feature fusion for malignancy characterization of hepatocellular carcinoma, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 472–479.
- [32] J. Yao, X. Zhu, F. Zhu, J. Huang, Deep correlational learning for survival prediction from multi-modality data, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 406–414.
- [33] Y. Panagakis, M.A. Nicolaou, S. Zafeiriou, M. Pantic, Robust correlated and individual component analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (8) (2015) 1665–1678.
- [34] Z. Wang, J. Lu, R. Lin, J. Feng, J. Zhou, Correlated and individual multi-modal deep learning for RGB-D object recognition, *CoRR* (Apr. 2016) 4321–4330 vol. arXiv:arXiv:1604.01655.
- [35] U. Meier, A note on the power of Fisher's least significant difference procedure, *Pharm. Stat.* 5 (4) (2010) 253–263.
- [36] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A Gomez, L Kaiser, I Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [37] K. He, X. Zhang, J.S. S.Ren, Deep Residual Learning for Image Recognition, in: *Proceedings of the Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [38] M. Gour, S. Jain, T. Sunil Kumar, Residual learning based CNN for breast cancer histopathological image classification, *Int. J. Imaging Syst. Technol.* 30 (3) (2020) 621–635.
- [39] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, in: *Proceedings of the International Conference on Machine Learning*, PMLR, 2019, pp. 6105–6114.
- [40] L. Zhu, Q. She, D. Li, et al., Unifying nonlocal blocks for neural networks, *IEEE/CVF International Conference on Computer Vision* (2021) 12292–12301.
- [41] H. Hermessi, O. Mourali, E. Zagrouba, Multimodal medical image fusion review: theoretical background and recent advances, *Signal Process.* 183 (2021) 108036.