Computers in Biology and Medicine 107 (2019) 47–57

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Grading of hepatocellular carcinoma using 3D SE-DenseNet in dynamic enhanced MR images

Qing Zhou[a,b], Zhiyong Zhou[b], Chunmiao Chen[c], Guohua Fan[d], Guangqiang Chen[d], Haiyan Heng[d], Jiansong Ji[c,**], Yakang Dai[b,*]

[a] University of Science and Technology of China, Hefei, 230026, China
[b] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, 215163, China
[c] Key Laboratory of Imaging Diagnosis and Minimally Invasive Intervention Research, Affiliated Lishui Hospital of Zhejiang University, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui Central Hospital, Lishui, 323000, China
[d] Department of Radiology, The Second Affiliated Hospital of Soochow University, Suzhou, 215004, China

ARTICLE INFO

Keywords:
3D DenseNet
SENet
Hepatocellular carcinoma
Grading
MRI

ABSTRACT

Background: Clinical histological grading of hepatocellular carcinoma (HCC) differentiation is of great significance in clinical diagnoses, treatments, and prognoses. However, it is challenging for radiologists to evaluate HCC gradings from medical images.
Purpose: In this study, a novel deep neural network was developed by combining the squeeze-and-excitation networks (SENets) in a three-dimensional (3D) densely connected convolutional network (DenseNet), which is referred to as a 3D SE-DenseNet, for the classification of HCC grading using enhanced clinical magnetic resonance (MR) images obtained from two different clinical centers.
Method: In the proposed architecture, the SENet was added as an additional layer between the dense blocks of the 3D DenseNet, to mitigate the impact of feature redundancy. For the HCC grading task, the 3D SE-DenseNet was trained after data augmentation, and it outperformed the 3D DenseNet based on the clinical dataset.
Results: The quantitative evaluations of the 3D SE-DenseNet on a two-class HCC grading task were conducted based on the dataset, which included 213 samples of the dynamic enhanced MR images. The proposed 3D SE-DenseNet demonstrated an accuracy of 83%, when compared with the 72% accuracy of the 3D DenseNet.
Conclusion: Owing to the advantage of useful automatic feature learning by the SE layer, the 3D SE-DenseNet can simultaneously handle useful feature enhancement and superfluous feature suppression. The quantitative experiments confirm the excellent performance of the 3D SE-DenseNet in the evaluation of the HCC grading.

1. Introduction

Diagnosis and treatment in the early stages of hepatocellular carcinoma (HCC) play a key role in the improvement of the survival rate [1]. Clinically, HCC grading after operations is subjective and time-consuming. It is therefore necessary to non-invasively and automatically evaluate the HCC gradings from clinical images prior to operations. In the literature, although several image processing methods were proposed for the computer-aided diagnosis of liver disease using different imaging modalities such as liver fat quantification [2], automated HCC grading methods are insufficient.

In recent years, many machine learning models have been applied to lesion and tumor classification, wherein the feature extraction is a critical step. To realize the classification of tumors, the texture features [3–7], shape features [3], and deep learning features [3,8] are extracted as classifiers. The abovementioned hand-crafted features have been widely used in several applications. However, Xu et al. [9] compared the performances of the manual features and deep learning features for the classification of tumors, where deep learning features exhibited an excellent ability to describe the tumor characteristics, and outperformed the manual texture features.

Classifiers are a major part of tumor classification and grading, i.e., the artificial neural network (ANN) [4,10], support vector machine (SVM) [4,5,8,11–13], extreme learning machine (ELM) [6], random forests [7], and back propagation neural network (BPNN) [3]. The focus of several recent researches was on deep features from hybrid models

* Corresponding author. Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, 215163, China.
** Corresponding author.
E-mail addresses: lschrjjs@163.com (J. Ji), daiyk@sibet.ac.cn (Y. Dai).

https://doi.org/10.1016/j.compbiomed.2019.01.026
Received 19 October 2018; Received in revised form 30 January 2019; Accepted 30 January 2019

for tumor classification [14]. Wang et al. [15] presented a multi-model-based framework for HCC grading, which achieved an accuracy of 91%. This framework consisted of an AdaBoost classifier and BPNN classifiers, which were trained by the deep features from a multi-view (axial, coronal, and sagittal) convolutional neural network (CNN). Li et al. [1] proposed a composite architecture that combined a multiple fully-connected CNN with an ELM for HCC nuclei grading. Moreover, this method achieved an accuracy of 81.1%. These hybrid models are examples of decision-level fusion methods. However, the complexity of these multi-model-based methods limits their generalization and application to various tasks.

With the rapid development of deep learning, CNN has emerged as a popular approach for visual pattern recognition tasks [16]. Moreover, CNN-based methods demonstrate excellent performances in lesion grading tasks. Bevilacqua et al. [17] proposed a CNN framework trained by gray level co-occurrence matrix features for HCC grading, which achieved an average accuracy of 92.8%. Liu et al. [18] presented a multi-view CNN framework trained by a multi-channel input, which consisted of different view areas of lung nodules for lung nodule classification, and it achieved an accuracy of 94%. Ertosun et al. [19] proposed a CNN-based cascaded classification architecture for the realization of automated gliomas grading.

However, these CNN-based models for lesion classification and grading are not sufficiently deep; thus, deep features cannot be effectively extracted. Khawaldeh et al. [20] adopted an Alex Krizhevsky Network (AlexNet) with a significant depth increase for glioma tumor grading, which achieved an accuracy of 91% when applied to the MR images [20]. Recently, several feature fusion methods were proposed to solve the gradient vanishing problem caused by the depth increase of networks. Ishikawa et al. [21] adopted the Residual Network (ResNet) for brain tumor classification from brain cell images. Chang et al. [22] adopted the ResNet for glioma grading in MR images. He et al. [23] proposed the ResNet, which re-uses features through skip connection to overcome the gradient vanishing problem. However, the ResNet simple re-uses the features and cannot explore new features. At the conference on Computer Vision and Pattern Recognition (CVPR), 2017, Huang et al. [24] proposed the Densely Connected Convolutional Networks (DenseNet) framework, which demonstrated an excellent performance with respect to image classification based on the exploration of new features using the dense block. Dey et al. [25] made a comparison between the DenseNet, basic 3D CNN, and 3D multi-output CNN for the diagnostic classification of benign and malignant lung nodules; wherein the DenseNet achieved the highest accuracy of 90.4%.

However, accurate HCC grading in MR images based on the DenseNet remains a challenging task for the following reason. In the dense block, each layer is connected to every other layer in a feed-forward manner, which results in the redundancy of features [26]. To solve the problem, in this study, a framework termed the 3D SE-DenseNet for HCC grading in MR images was developed. The main contributions of this work are as follows. First, the Squeeze-and-Excitation Networks (SENet) [27] was combined with a 3D DenseNet. The SENet can effectively enhance useful feature channels and suppress less-useful feature channels for the HCC grading task, and therefore realize the adaptive re-calibration of channel-wise feature responses. The combination can mitigate the impact of feature redundancy and improve the accuracy by 11%, which slightly increases the training time and number of parameters. Second, few researchers are currently focused on the direct application of the deep learning framework to HCC grading based on medical images. The MR images were associated with the pathology findings based on the proposed deep learning framework. An investigation of this association is essential for the computer-aided study on image diagnoses.

For the HCC grading task, an Edmondson and Steiner system is the most popular HCC grading system, which divides the HCC into four grades (from I–IV) based on histological differentiation [28,29]. In addition to the four-grade system, the HCCs can also be divided into high grades and low grades, given that most of the HCCs appear as grade II or III. The high grade includes grades I and II, and the low grade includes grades III and IV [30].

The rest of the paper is organized as follows. In Section 2, the detailed experimental data and methods are presented, and the proposed model (3D SE-DenseNet) is described in Subsection 2.2. A discussion on the grading performance of the 3D SE-DenseNet when applied to MR images obtained from two clinical centers is presented in Section 3. Section 4 presents a discussion on the HCC grading performance, followed by the conclusions in Section 5.

2. Materials and methods

2.1. Dataset and preprocessing

2.1.1. Dataset

The HCC dataset of the MR images was obtained from The Central Hospital of Lishui and The Second Affiliated Hospital of Soochow University. The dataset included 75 patients (66 men and 9 women) who were subjected to liver partial hepatectomy or needle biopsy. In the dataset, 37 patients were diagnosed with low-grade HCC, and 38 patients were diagnosed with high-grade HCC. Further details on the tumor sizes and slice thicknesses are presented in Table 1.

2.1.2. Workflow of 3D SE-DenseNet for HCC grading

The proposed classification workflow is presented in Fig. 1. The workflow consists of three major steps. The first step is data pre-processing for the 3D SE-DenseNet. The second step is data augmentation, and the third step is the training and testing of the 3D SE-DenseNet.

2.1.3. Data preprocessing

The detailed data pre-processing procedure is described as follows:

(1) Manual tumor delineation: in the dataset, 63 patients have the MR images of enhanced arterial phase, enhanced venous phase and enhanced delayed phase. The remaining http://dict.youdao.com/w/remaining/javascript:void(0); 12 patients only have the MR images of enhanced arterial phase and enhanced venous phase. Hence, 213($63 \times 3 + 12 \times 2$) tumor samples were simply manually segmented. Examples of the tumors in the dynamic enhanced MR images are presented in Fig. 2.
(2) Three-dimensional tumor ROI extraction: the smallest cuboid that can cover the entire tumor was used. The voxels of the tumor in this cuboid remained the same as the original MR images, and the background voxels of the non-tumor were set to zero. The example images are presented in the middle column of Fig. 3.
(3) Normalization: The dimensions of the normalized image size were set as $200 \times 200 \times 10$ (voxel). The images with sizes less than the normalized size were padded with zero, and the remaining images with sizes larger than the normalized size were truncated, to guarantee that each tumor is located at the center of the normalized image. After the image size was normalized, the voxel value of the tumor was also normalized using the Z-score normalization method: subtraction of the mean and then division by the variance. The

Table 1
Details of HCC dataset.

Grading		Low	High
Tumor Size	≤3 cm	14	14
	3–6 cm	17	10
	6–9 cm	6	9
	> 9 cm	0	5
Slice Thickness	8 mm	23	24
	6 mm	8	4
	3 mm	6	10

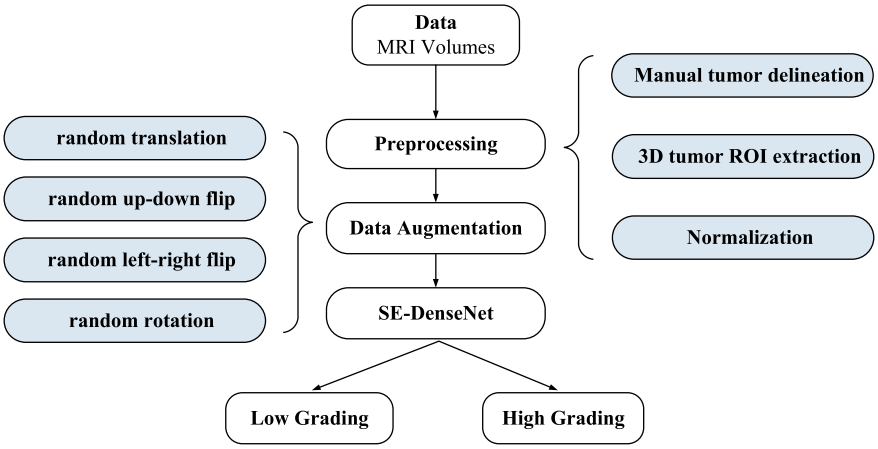

Fig. 1. Workflow of HCC grading classification.

example images are presented in the right column of Fig. 3.

2.1.4. Data augmentation

After the data pre-processing, 213 tumor ROIs, which consisted of 75 enhanced arterial phase tumors, 75 enhanced venous phase tumors, and 63 enhanced delayed phase tumors were made available. Moreover, 80% of all the tumor ROIs were used as training data, and the remaining 20% were used as testing data. To increase the number of training data, several data augmentation methods were adopted. In particular, each ROI was operated by random translation, random up-down flip, random left-right flip, and random rotation at a random angle. With these operations, the number of training data was augmented by a factor of 15. Finally, a total of 2550 training ROIs and 43 testing ROIs were inputted to the 3D SE-DenseNet for the HCC grading task.

2.2. The proposed model: 3D SE-DenseNet

2.2.1. 3D SE-DenseNet

The proposed framework, as illustrated in Fig. 4, mainly consists of three dense blocks, two transition layers, and two SE layers. The dense block is made up of 12 composite functions. The composite function illustrated in Fig. 5 is defined by four consecutive operations: a batch normalization (BN), a rectified linear unit (ReLU), a $3 \times 3 \times 3$ convolution layer, and a dropout layer. The BN reduces the internal covariate shift, which can speed up the training process and increase the accuracy [31]. The ReLU reduced the probability of the vanishing gradient [31]. The dropout layer can randomly reduce a branch to avoid over-fitting [31]. The transition layer illustrated in Fig. 4 is set between two dense blocks. As shown in Fig. 6, the transition layer consists of a BN, a ReLU, a $1 \times 1 \times 1$ convolution layer, a dropout layer, and a $2 \times 2 \times 2$ average pooling layer. As illustrated in Fig. 7, the SE layer is made up of five consecutive operations: a global average pooling layer, a fully connected layer, a ReLU, a fully connected layer, and a sigmoid. The final output of the SE layer is obtained by multiplying the input channels with respective weights. In the SE layer, there is a hyper-parameter ratio, namely, the channel reduction ratio, which enhances the most important 1/ratio percentage of the feature channels.

The details of the proposed framework are presented in Table 2. As can be seen in Table 2, the images are inputted to a 3D convolution layer with kernel dimensions of $7 \times 7 \times 7$. The following layer is a $3 \times 3 \times 3$ average pooling layer that significantly reduces the computational load of the training process. The number of output channels for the convolution layer is twice the growth rate. The hyper-parameter growth rate is the number of output channels for the composite function. After the first convolution layer and pooling layer, the following sections are the dense block 1, transition layer 1, SE layer 1, dense block

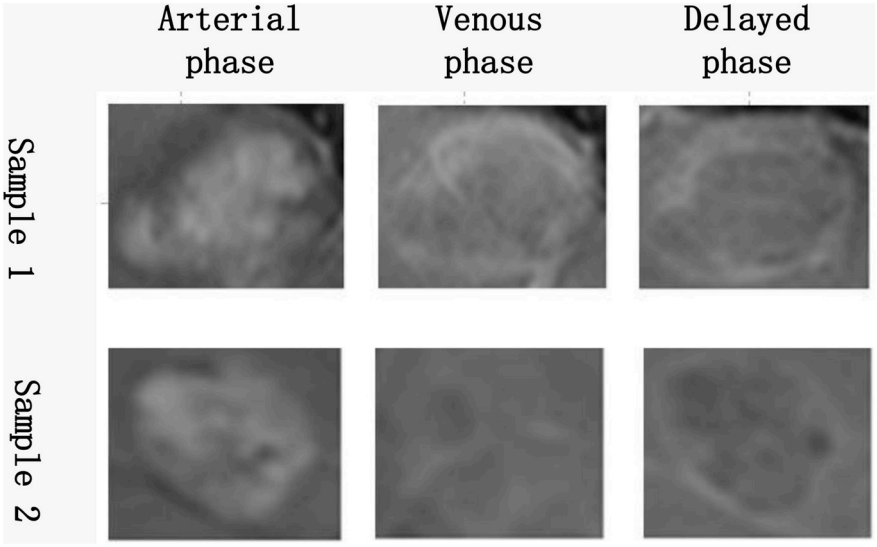

Fig. 2. HCC tumor MR image patches in the experiment. The three images in the first row are from the same patient diagnosed with low-grade HCC. The three images in the second row are from the same patient diagnosed with high-grade HCC. Each image presented in this figure is one slice of a 3D image.

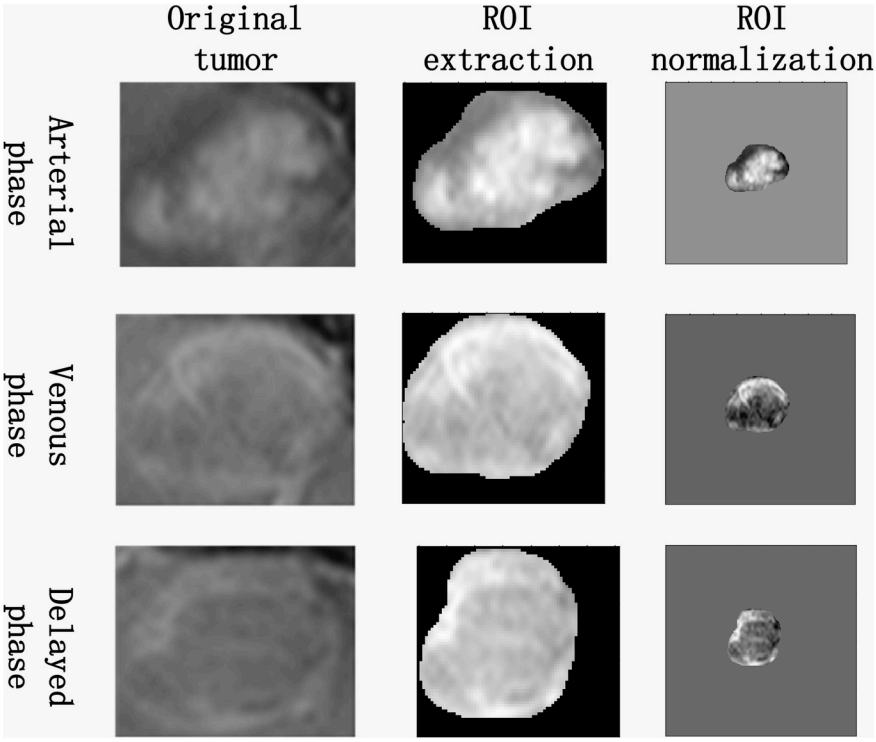

Fig. 3. Pre-processing step: the three rows present MR images of the arterial phase, venous phase, and delayed phase, respectively. The images in the first column display tumors from the original images. The images in the second column display tumor ROIs extracted from the original images. The images in the third column are the tumor regions of interest (ROIs) obtained from the normalization step (After the Z-score normalization step, the value of the normalized image voxel is very small. The value was mapped to the range of 0–255 for the display).

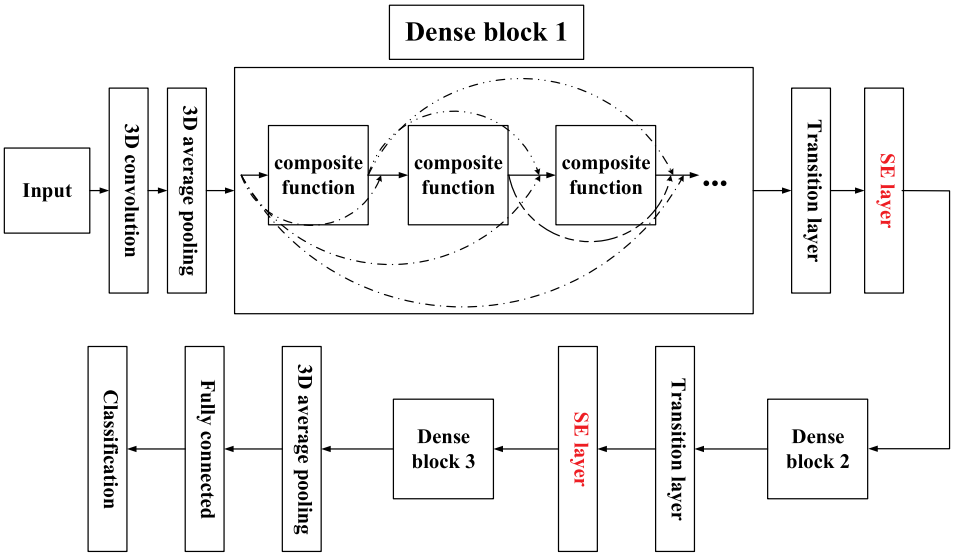

Fig. 4. Framework of 3D-SE-DenseNet.

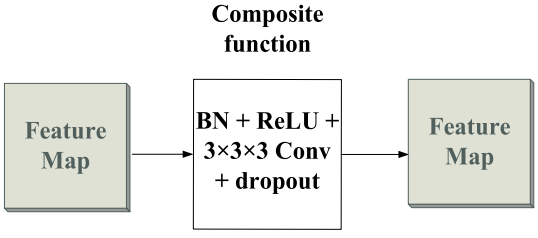

Fig. 5. Detailed structure of composite function.

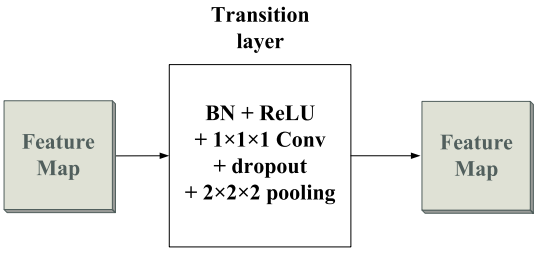

Fig. 6. Detailed structure of transition layer.

2, transition layer 2, SE layer 2, and dense block 3, successively. Finally, the feature maps are inputted to the classification layer, which consists of a global average pooling layer, a fully connected layer, and a softmax classifier.

2.2.2. The variants of 3D SE-DenseNet

In this paper, the following three methods for combining the SENet with a 3D DenseNet are proposed. (1) The SE layer is set between the

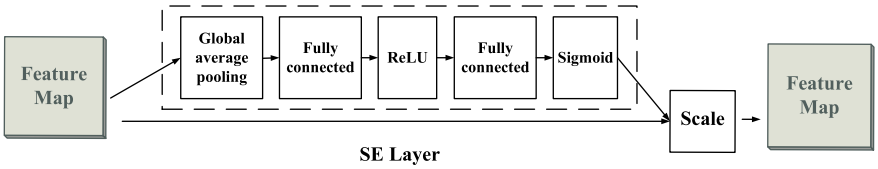

Fig. 7. Detailed structure of SE layer.

Table 2
The architecture of the 3D SE-DenseNet. The growth rate described in the context is set as 24. Each "Conv." layer and "scale" layer presented in the table correspond to the BN–ReLU–3DConvolution–dropout sequence and the global average pooling–fully connected–ReLU-fully-connected–sigmoid sequence (illustrated in Fig. 7), respectively.

Layers	Output Size	3D SE-DenseNet (growth rate = 24)
3D Convolution	10 × 50 × 50	7 × 7 × 7 Conv. with stride 1 × 2 × 2
Dense Block 1	10 × 50 × 50	[3 × 3 × 3 Conv.] × 12
Transition Layer 1	10 × 50 × 50	1 × 1 × 1 Conv.
	5 × 25 × 25	2 × 2 × 2 average pool
SE layer 1	5 × 25 × 25	scale × 1
Dense Block 2	5 × 25 × 25	[3 × 3 × 3 Conv.] × 12
Transition Layer 2	5 × 25 × 25	1 × 1 × 1 Conv.
	3 × 13 × 13	2 × 2 × 2 average pool
SE layer 2	3 × 13 × 13	scale × 1
Dense Block 3	3 × 13 × 13	[3 × 3 × 3 Conv.] × 12
Classification Layer	1 × 1 × 1	13 × 13 × 13 average pooling
	–	fully connected
	–	classification and softmax

dense blocks, which is referred to as a 3D SE-DenseNet, as illustrated in Fig. 9(a); (2) the SE layer is set within the dense block, which is referred to as a 3D SE-DenseNet-IN, as illustrated in Fig. 9(b); or (3) the SE layer is set between and within the dense blocks, which is referred to as a 3D SE-DenseNet-ALL, as illustrated in Fig. 9(c). The dense block is illustrated in Fig. 8(a); and the SE-dense block, wherein the SE layer is set within the dense block, is illustrated in Fig. 8(b). In addition, the 3D DenseNet-BC is also combined with the SENet in three such manners, as follows. (1) The SE layer is set between the dense-BC blocks, which is referred to as a 3D SE-DenseNet-BC-BE, as illustrated in Fig. 9(d); (2) the SE layer is set within the dense-BC block, which is referred to as a 3D SE-DenseNet-BC-IN, as illustrated in Fig. 9(e); and (3) the SE layer is set between and within the dense-BC blocks, which is referred to as a 3D SE-DenseNet-BC-ALL, as illustrated in Fig. 9(f). The dense-BC block is illustrated in Fig. 8(c); and the SE-dense-BC block, wherein the SE layer is set within the dense-BC block, is illustrated in Fig. 8(d).

3. Results

All the deep learning networks referred to in the results section were implemented using a Python tool in the Tensorflow framework. The computations were run on an Intel Xeon E31220 3.1 GHz PC with 12 GB of random-access memory (RAM) and a NVIDIA GeForce GTX 1080 Ti graphics processing unit (GPU).

3.1. Parameter initialization

To adjust all the 3D tumor ROIs to the same size, the size distributions of

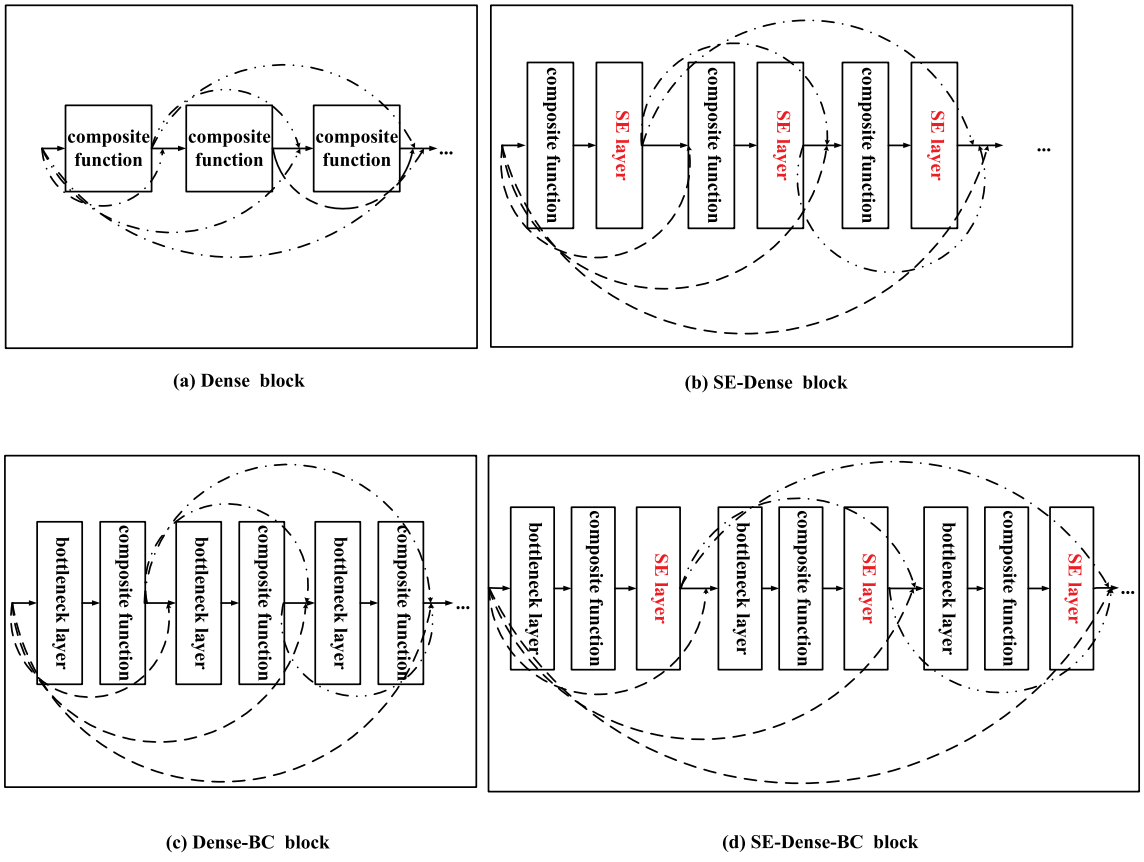

Fig. 8. The variants of the dense block: (a) dense block, (b) SE-dense block, (c) dense-BC block, and (d) SE-dense-BC block.

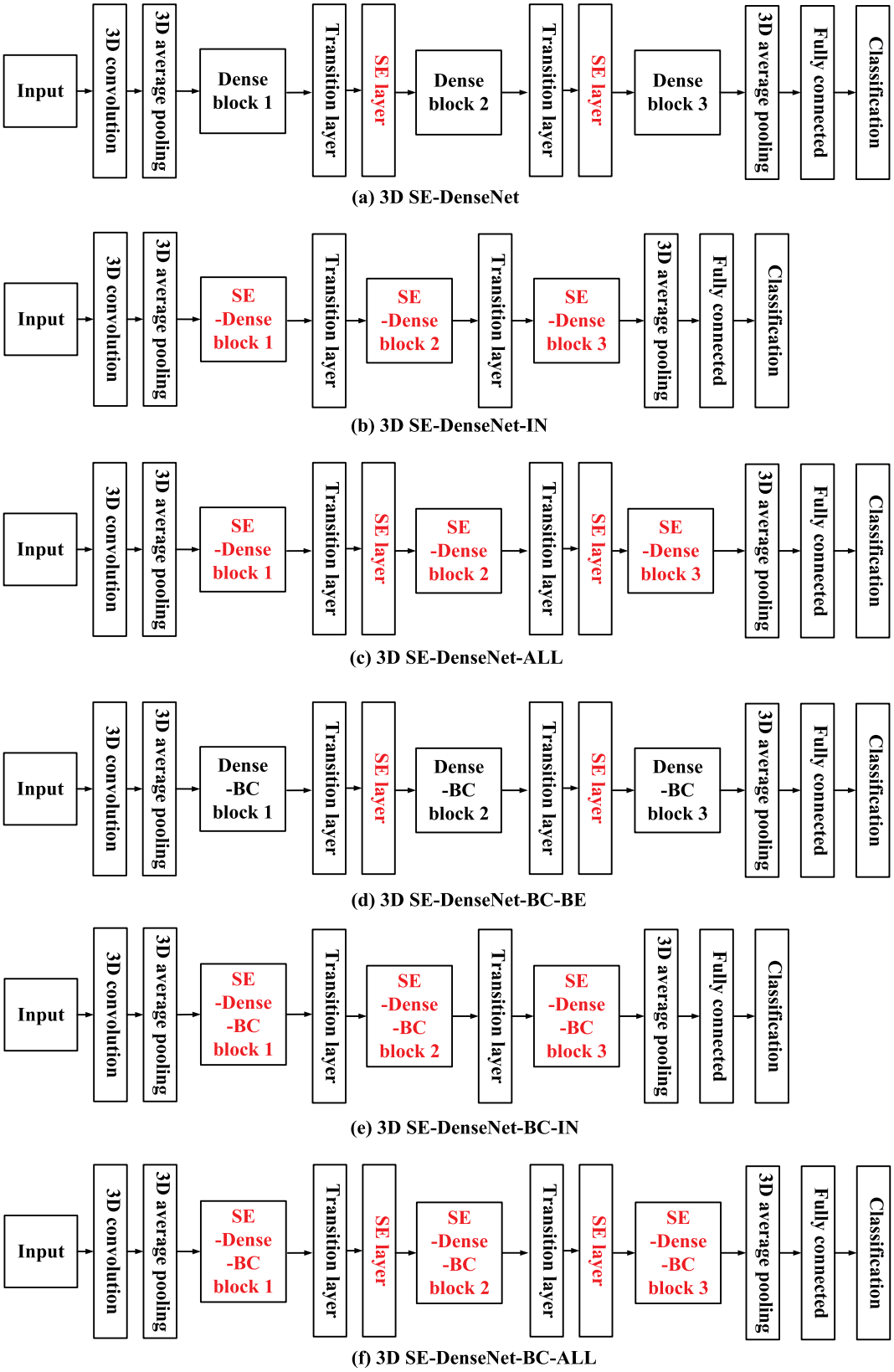

Fig. 9. Variants of 3D SE-DenseNet: (a) dense block, (b, c) SE-dense block, (d) dense-BC block, and (e, f) SE-dense-BC block; as illustrated in Fig. 8(a), (b), (c), and (d), respectively.

all the 3D tumor ROIs were analyzed in the axial, coronal, and sagittal directions. The statistical results, as shown in Fig. 10, reveal that the dimensions of the tumors were mostly less than 256 × 256 × 20 (voxel). To include the tumor voxels: in this study, the following normalized image dimensions were set: 256 × 256 × 20, 256 × 256 × 10, 160 × 160 × 40, and 200 × 200 × 10. With the consideration of the hardware

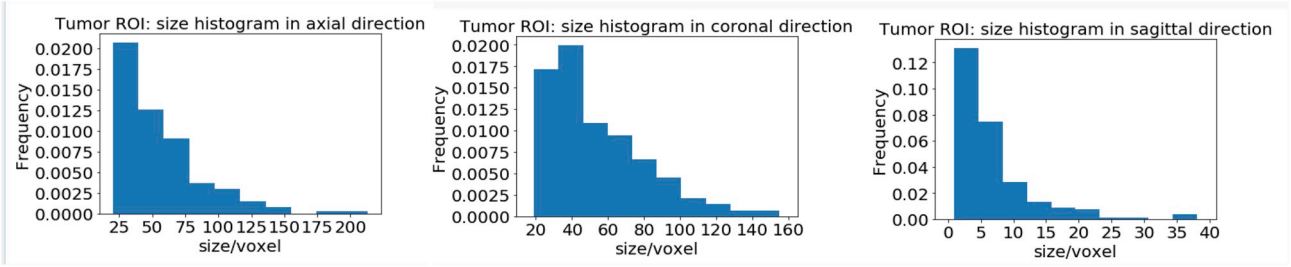

Fig. 10. Distribution of tumor size in axial, coronal, and sagittal directions.

Table 3
Basic parameters of 3D SE-DenseNet.

Hyper-parameters	Value
Image size/Batch size	(200, 200, 10)/10
Keep probability	0.5
Depth/weight decay	20/0.0001; 30/0.0001; 40/0.0001; 50/0.00001
Growth rate	24
Ratio of the SENet	2

configuration, the batch size was set as 2, 4, 2, and 10 maximumly. Within a suitable range, a relatively large batch size results in a good performance, especially with respect to the training time and convergence precision. If the dataset is sufficient, the batch size can be set between 1 and several hundreds. For example, 32 is a good default value [32]. By comparing the training curves (training and validation error with respect to the training time), the trade-off between the model performance and hardware configuration can be realized when the normalized image size and batch size of the model are set as $200 \times 200 \times 10$ and 10, respectively (normalized image size/accuracy: $256 \times 256 \times 20$/69%, $256 \times 256 \times 10$/75%, $160 \times 160 \times 40$/71%, $200 \times 200 \times 10$/83%).

A stochastic gradient [33] was employed as the optimizer in the training procedure. Given the limitation of the RAM and GPU memory, the number of training data was augmented to a number greater than the original number by a factor of 15, the dimensions of the ROI (width × height × length) were set as $200 \times 200 \times 10$, and the respective batch size was set as 10. The growth rate, depth of the model, and ratio of the SE layer are presented in Table 3. In the evaluations, the growth rate was limited to a small integer such as 24, to prevent the excessively rapid growth of the 3D DenseNet [31]. Moreover, the increase in the sample sensitivity and parameter number with an increase in the depth may cause over-fitting. In the experiments, the results at different depths (20, 30, 40, and 50) were compared. With the exception of the hyper-parameters, the keep probability in dropout layer was set as 0.5. The value of the weight decay was determined according to a rule: the weight decay multiplying L_2 Regularization Item loss should be in the same order of magnitude as the cross entropy. The training epoch was set as 70 and the initial learning rate was 0.1. A decrease in the initial learning rate to 0.01 from 35th epoch and 0.001 from 52nd epoch was expected.

3.2. Performance of 3D SE-DenseNet

A comparison was made between the proposed 3D SE-DenseNet, 3D DenseNet [24], 3D ResNet [22], and 3D AlexNet [20] with respect to several parameters, which included the recall, precision, area under curve (AUC), F1-score, Matthews correlation coefficient (MCC), geometric mean (G-Mean), and accuracy. Moreover, 10 five-fold cross-validation procedures were carried out. In particular, 20% of all the tumor ROIs (total ROI count of 213) were used as testing samples, and the remaining 80% were used as training samples in each cross-validation procedure.

A comparison of the cross-validation results for the SE-DenseNet, DenseNet, ResNet, and AlexNet is presented in Table 4. The 3D SE-DenseNet (83%) and 3D DenseNet (72%) both outperformed the 3D ResNet (70%) and 3D AlexNet (56%). Moreover, the 3D SE-DenseNet achieved an accuracy of 83%, which exceeded that of the 3D DenseNet (72%) by 11%. The clear improvement indicates that the 3D SE-DenseNet can increase the parameter efficiency and reduce over-fitting.

Magnetic resonance images generally contain noise and several artifacts. In the data pre-processing step, there is no noise reduction operation. Therefore, the data inputted to the SE-DenseNet contains noise. Moreover, to quantitatively evaluate the performance of the method in the presence of noise, Gaussian noise and pepper-and-salt noise were added to the original data. With Gaussian noise, the parameters of the mean and variance were 0 and 0.01, respectively. With the pepper-and-salt noise, the proportion of the image pixels replaced with noise was set as 0.05, and the proportion of salt noise with respect to pepper noise was set as 0.5. The comparison of the results between the original data and noisy data is presented in Table 5. The small decrease in accuracy (4% for Gaussian noise and 4% for the pepper-and-salt noise when compared with the original data) indicates that the proposed 3D SE-DenseNet is relatively stable when subject to data that contains noise.

3.3. Comparison between different variants of 3D SE-DenseNet

In this subsection, several cross-validation procedures were considered to compare the HCC grading performances of variant models derived from the 3D DenseNet. All the comparison results are presented in Table 6. As can be seen in Table 6, the image classification performances and training costs of all the models were evaluated. With respect to the training cost, the focus was on the training speed (running time) and memory cost (depth and parameters). The standard for the image classification performance included the recall, precision, AUC,

Table 4
Results of the comparison between the 3D SE-DenseNet, 3D DenseNet, 3D ResNet, and 3D AlexNet. Each result is reported in the form of mean ± standard deviation. The bold is the line with the highest accuracy.

Model	Recall	Precision	AUC	F1-score	MCC	G-Mean	Accuracy
SE-DenseNet	**0.79 ± 0.034**	**0.80 ± 0.033**	**0.83 ± 0.029**	**0.79 ± 0.036**	**0.59 ± 0.069**	**0.78 ± 0.047**	**0.83 ± 0.028**
DenseNet [24]	0.68 ± 0.067	0.73 ± 0.028	0.79 ± 0.032	0.66 ± 0.092	0.41 ± 0.075	0.64 ± 0.105	0.72 ± 0.049
ResNet [22]	0.67 ± 0.048	0.68 ± 0.048	0.74 ± 0.050	0.66 ± 0.053	0.35 ± 0.099	0.65 ± 0.063	0.70 ± 0.052
AlexNet [20]	0.55 ± 0.043	0.61 ± 0.067	0.58 ± 0.052	0.50 ± 0.088	0.15 ± 0.071	0.43 ± 0.153	0.56 ± 0.044

Table 5
Performance comparison results for 3D SE-DenseNet using data that contains noise. Each result is reported in the form of mean ± standard deviation.

Input	Recall	Precision	AUC	F1-score	MCC	G-Mean	Accuracy
Original data	0.79 ± 0.034	0.80 ± 0.033	0.83 ± 0.029	0.79 ± 0.036	0.59 ± 0.069	0.78 ± 0.047	0.83 ± 0.028
Original data + Gaussian	0.75 ± 0.031	0.75 ± 0.030	0.80 ± 0.027	0.75 ± 0.032	0.51 ± 0.062	0.75 ± 0.034	0.79 ± 0.026
Original data + Pepper and Salt	0.76 ± 0.038	0.77 ± 0.049	0.82 ± 0.050	0.76 ± 0.036	0.54 ± 0.086	0.76 ± 0.036	0.79 ± 0.031

Table 6
Results of the comparison between all the architectures at different depths. Sub-table (a) is for a depth of 40, and Sub-table (b) is for a depth of 30. Each result is reported in the form of mean ± standard deviation. The bold is the line with the highest accuracy.

Model	Running time	Parameters	Recall	Precision	AUC	F1-score	MCC	G-Mean	Accuracy
(a)									
DenseNet	6h16	11.5M	0.68 ± 0.067	0.73 ± 0.028	0.79 ± 0.032	0.66 ± 0.092	0.41 ± 0.075	0.64 ± 0.105	0.72 ± 0.049
SE-DenseNet	**6h25**	**12.8M**	**0.79 ± 0.034**	**0.80 ± 0.033**	**0.83 ± 0.029**	**0.79 ± 0.036**	**0.59 ± 0.069**	**0.78 ± 0.047**	**0.83 ± 0.028**
SE-DenseNet-IN	6h37	11.5M	0.73 ± 0.035	0.73 ± 0.038	0.76 ± 0.057	0.73 ± 0.035	0.46 ± 0.073	0.73 ± 0.034	0.75 ± 0.059
SE-DenseNet-ALL	6h39	12.8M	0.73 ± 0.086	0.74 ± 0.093	0.77 ± 0.085	0.73 ± 0.088	0.46 ± 0.177	0.72 ± 0.092	0.77 ± 0.069
DenseNet-BC	2h43	1.4M	0.64 ± 0.065	0.69 ± 0.055	0.69 ± 0.090	0.61 ± 0.097	0.31 ± 0.105	0.56 ± 0.136	0.66 ± 0.090
SE-DenseNet-BC-BE	2h44	1.5M	0.68 ± 0.056	0.70 ± 0.051	0.76 ± 0.048	0.66 ± 0.079	0.36 ± 0.122	0.62 ± 0.116	0.68 ± 0.103
SE-DenseNet-BC-IN	2h53	1.5M	0.67 ± 0.063	0.73 ± 0.054	0.74 ± 0.073	0.64 ± 0.096	0.38 ± 0.116	0.60 ± 0.154	0.67 ± 0.104
SE-DenseNet-BC-ALL	2h54	1.6M	0.69 ± 0.076	0.72 ± 0.078	0.78 ± 0.061	0.67 ± 0.093	0.38 ± 0.153	0.63 ± 0.127	0.71 ± 0.098
(b)									
DenseNet	3h16	5.3M	0.66 ± 0.041	0.71 ± 0.041	0.78 ± 0.034	0.64 ± 0.064	0.36 ± 0.069	0.61 ± 0.088	0.69 ± 0.082
SE-DenseNet	**3h20**	**6.0M**	**0.76 ± 0.047**	**0.77 ± 0.034**	**0.81 ± 0.046**	**0.75 ± 0.053**	**0.53 ± 0.068**	**0.75 ± 0.053**	**0.80 ± 0.040**
SE-DenseNet-IN	3h30	5.3M	0.71 ± 0.044	0.72 ± 0.049	0.76 ± 0.050	0.71 ± 0.044	0.43 ± 0.093	0.70 ± 0.044	0.74 ± 0.045
SE-DenseNet-ALL	3h34	6.0M	0.73 ± 0.040	0.73 ± 0.045	0.77 ± 0.038	0.73 ± 0.040	0.46 ± 0.087	0.73 ± 0.044	0.77 ± 0.036
DenseNet-BC	1h49	0.9M	0.59 ± 0.062	0.73 ± 0.054	0.66 ± 0.065	0.50 ± 0.109	0.23 ± 0.071	0.39 ± 0.172	0.61 ± 0.085
SE-DenseNet-BC-BE	1h50	1.0M	0.70 ± 0.065	0.74 ± 0.041	0.80 ± 0.037	0.68 ± 0.093	0.42 ± 0.107	0.63 ± 0.132	0.71 ± 0.081
SE-DenseNet-BC-IN	1h56	1.0M	0.63 ± 0.075	0.66 ± 0.078	0.71 ± 0.087	0.62 ± 0.082	0.29 ± 0.137	0.60 ± 0.076	0.64 ± 0.115
SE-DenseNet-BC-ALL	1h58	1.0M	0.69 ± 0.057	0.72 ± 0.066	0.78 ± 0.059	0.67 ± 0.075	0.39 ± 0.130	0.64 ± 0.107	0.73 ± 0.067

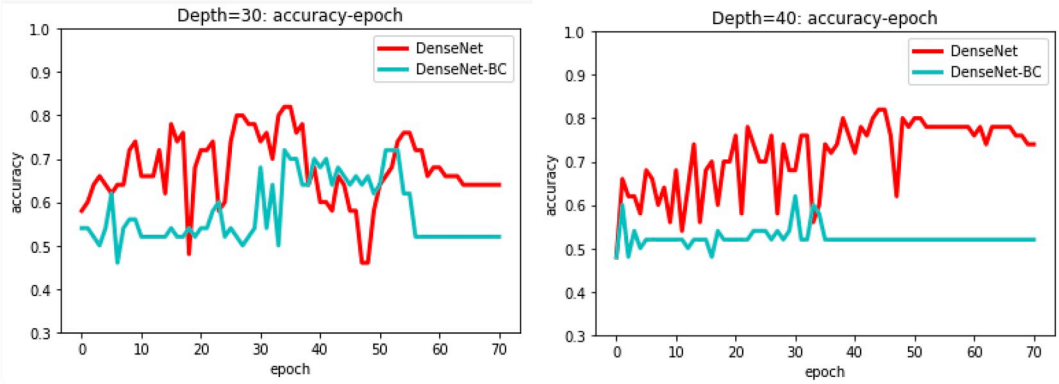

Fig. 11. Accuracy of 3D DenseNet and 3D DenseNet-BC at different depths. The figure on the left is for a depth of 30, and the figure on the right one is for a depth of 40. The curves of the accuracy with respect to the training epoch are described.

F1-score, MCC, G-Mean, and accuracy.

The results of the 3D DenseNet and 3D DenseNet-BC were compared, as presented in Table 6 and Fig. 11. Overall, the accuracy of the 3D DenseNet was superior to that of the 3D DenseNet-BC (72% and 66% at a depth of 40; and 69% and 61% at a depth of 30, respectively), despite the higher training cost. Therefore, for the HCC grading task, the 3D DenseNet is more appropriate.

As shown in Table 6 and Fig. 12, given that the SENet is added to the DenseNet, the stability and convergence of the improved DenseNet were superior to those of the DenseNet at the expense of a small increase in the training time and number of parameters. With the combination of the SENet and the DenseNet, changes in the image classification performance and the training cost were observed, as follows. (1) The increase in accuracy ranged from 3 to 11% at a depth of 40, and from 5 to 11% at a depth of 30; (2) the number of parameters maximally increased from 11.5M to 12.8M at a depth of 40, and from 5.3M to 6.0M at a depth of 30, respectively; and (3) the increase rate of the training time was less than 10%. In general, the relatively small increase in the training time and memory cost indicates that the performance improvement induced by the SE layer was computationally efficient and resulted in an excellent performance. Similarly, for the combination of the SENet with the DenseNet-BC, the performance of the improved DenseNet-BC for HCC grading was improved at the cost of an increase in the training time and the number of parameters. The abovementioned experimental results indicate that the SENet improves the performance of the 3D DenseNet and 3D DenseNet-BC. In summary, the novel scheme presented in this paper is based on the 3D DenseNet with respect to accuracy.

To evaluate the best method of combining the SENet with a 3D DenseNet, the performance of the 3D SE-DenseNet was compared with those of its variants, namely, the 3D SE-DenseNet-IN and 3D SE-DenseNet-ALL, as shown in Table 6. Moreover, the curves of the accuracy with respect to the training epoch are further presented in Fig. 13. From the experimental results, the accuracy of the 3D SE-DenseNet was found to be the highest among the three models, namely, the 3D SE-DenseNet, 3D SE-DenseNet-IN, and 3D SE-DenseNet-ALL

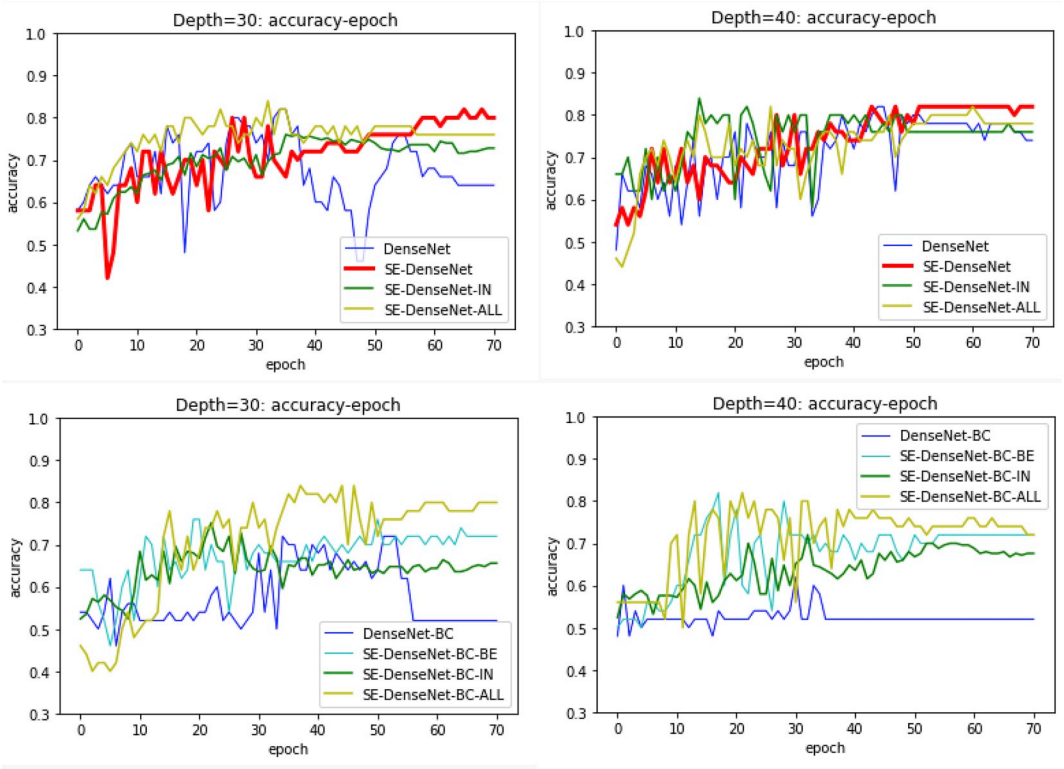

Fig. 12. Accuracy of different models. The two graphs in the first row present accuracy curves of the 3D DenseNet, 3D SE-DenseNet, 3D SE-DenseNet-IN, and 3D SE-DenseNet-ALL at different depths: the graph on the left is for a depth of 30, and the graph on the right is for a depth of 40. The two graphs in the second row present accuracy curves of the 3D DenseNet-BC, 3D SE-DenseNet-BC-BE, 3D SE-DenseNet-BC-IN, and 3D SE-DenseNet-BC-ALL at different depths: the graph on the left one is for a depth of 30, and the graph on the right is for a depth of 40. The curves of the accuracy with respect to the training epoch confirm the superior performance of the SENet.

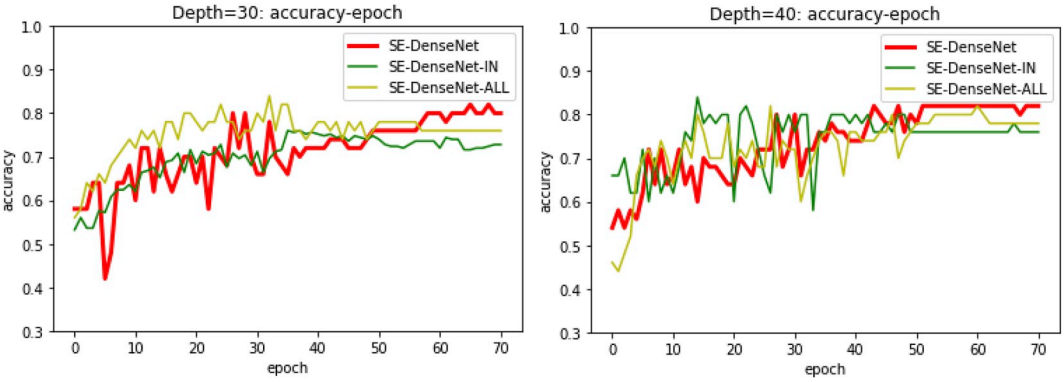

Fig. 13. Accuracy of the 3D SE-DenseNet, 3D SE-DenseNet-IN, and 3D SE-DenseNet-ALL at different depths. The graph on the left is for a depth of 30, and the graph on the right is for a depth of 40. The curves of the accuracy with respect to the training epoch reveal that the performance of the SE-DenseNet was the best among all the other models.

Table 7
Comparison of **3D SE-DenseNet** performance at different depths. Each result is reported in the form of mean ± standard deviation. The bold is the line with the highest accuracy.

Depth	Running time	Parameters	Recall	Precision	AUC	F1-score	MCC	G-Mean	Accuracy
20	1h46	2.5M	0.75 ± 0.043	0.75 ± 0.049	0.81 ± 0.052	0.74 ± 0.043	0.50 ± 0.095	0.74 ± 0.048	0.78 ± 0.036
30	3h20	6.0M	0.76 ± 0.047	0.77 ± 0.034	0.81 ± 0.046	0.75 ± 0.053	0.53 ± 0.068	0.75 ± 0.053	0.80 ± 0.040
40	**6h25**	**12.8M**	**0.79 ± 0.034**	**0.80 ± 0.033**	**0.83 ± 0.029**	**0.79 ± 0.036**	**0.59 ± 0.069**	**0.78 ± 0.047**	**0.83 ± 0.028**
50	10h5	19.7M	0.72 ± 0.054	0.72 ± 0.055	0.76 ± 0.074	0.71 ± 0.055	0.44 ± 0.113	0.71 ± 0.060	0.74 ± 0.052

(83%, 75%, and 77% at a depth of 40; and 80%, 74%, and 77% at a depth of 30, respectively). In addition, the number of parameters of the 3D SE-DenseNet and 3D SE-DenseNet-ALL were similar; however, the performance of the 3D SE-DenseNet was superior to that of the 3D SE-DenseNet-ALL with less training time. Compared with the 3D SE-DenseNet-IN, the performance of the 3D SE-DenseNet was significantly

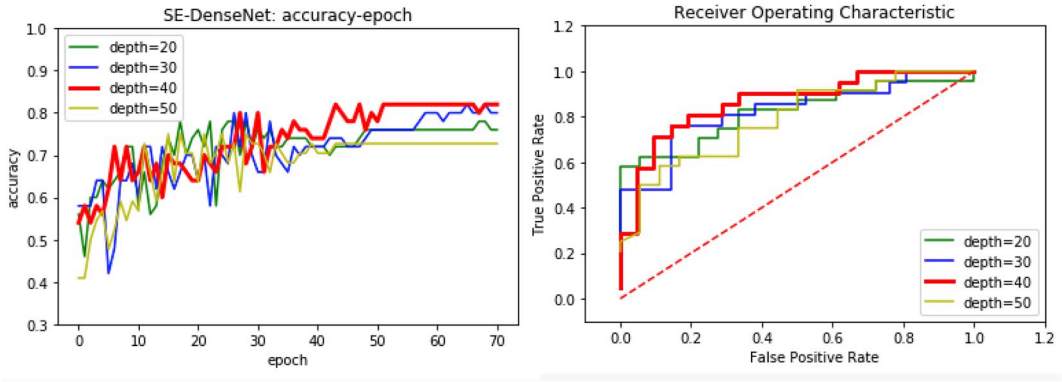

Fig. 14. The graph on the left presents the accuracy of the 3D SE-DenseNet with respect to the training epoch at different depths. The graph on the right presents the receiver operating characteristic (ROC) curves of the 3D SE-DenseNet at different depths. The red curve at a depth of 40 represents the best performance.

better, at the expense of a small increase in the training time and the number of parameters. With respect to the AUC, the stability of the 3D SE-DenseNet was the best among the three models, as confirmed by Fig. 13. The experimental results indicate that the placement of the SE layer between the dense blocks is more advantageous than that within the dense block. The proposed framework therefore employs the method of adding the SE layer between the dense blocks.

Moreover, the results of the 3D SE-DenseNet at different depths were compared, as shown in Table 7 and Fig. 14. The accuracy of 3D SE-DenseNet was 78% at a depth of 20, 80% at a depth of 30, 83% at a depth of 40, and 74% at a depth of 50. The number of parameters was 2.5M at a depth of 20, 6.0M at a depth of 30, 12.8M at a depth of 40, and 19.7M at a depth of 50. The depth of the proposed framework was then set as 40, to obtain the best performance for HCC grading.

4. Discussion

In the proposed 3D SE-DenseNet architecture for HCC grading tasks in dynamic enhanced MR imaging, the SENet was set between the dense blocks of the DenseNet. In the dense block, each layer is connected to all the other layers in a feed-forward manner, which allows for the extraction of new features based on the features of previous layers. However, the feature redundancy in the process of feature fusion is a major issue. Due to the advantage of useful automatic feature learning in the SENet, the 3D SE-DenseNet can suppress superfluous features, which mitigates the impact of feature redundancy to a certain extent and improves the performance for HCC grading tasks. The total training process of the 3D SE-DenseNet required 375 min for 70 epochs.

Based on the experimental results, the performance of 3D DenseNet is superior to that of the 3D DenseNet-BC for HCC grading tasks. Moreover, the feature dimension reduction in the 3D DenseNet-BC may not sufficiently and effectively extract the characteristics of the HCC differentiation. In addition, the benefits of the SENet both in the 3D DenseNet and 3D DenseNet-BC were confirmed. In section 3.2, the superior accuracy of the 3D SE-DenseNet over the 3D DenseNet was confirmed with respect to the HCC grading performance, at the cost of a small increase in the training time and memory. In addition to the DenseNet, the proposed SE-DenseNet outperformed the ResNet and the AlexNet, as described in Table 4. However, as can be seen in Fig. 13, the advantage of positioning the SE layer within the dense block was not as significant as that between the dense blocks. The SE layer within the dense block takes a weight learning of features in the procedure of exploring new features, which may cause over-fitting, and the SE layer between the dense blocks can enhance useful features and suppress superfluous features for HCC grading after the extraction of new features; which may have led to the more significant improvement of the 3D SE-DenseNet performance when compared with that of the 3D SE-DenseNet-IN.

In this study, there were two limitations. First, the quantity and quality of the dataset obtained from both clinic centers were limited. As can be seen in Table 1, the slice thicknesses of the most of the data were larger than 5 mm, which may cause a loss in the representation information of tumors. However, the proposed model can be applied to low-quality clinical images. Therefore, in future work, more MR images of the HCC patients should be collected. Second, the tumor ROIs were manually drawn, which is a time-consuming task, especially with respect to 3D medical images.

5. Conclusion

In this paper, a new model termed the 3D SE-DenseNet was proposed for HCC grading tasks in MR images. The experimental results reveal that the proposed model demonstrates a better performance than those of the 3D DenseNet and 3D DenseNet-BC. Moreover, it was confirmed that the SE layer helps fuse feature maps and improves the HCC grading performance of the 3D DenseNet. In addition, the proposed model performed well when applied to clinical MR images with slice thicknesses larger than 5 mm. It is therefore feasible to classify the histological HCC grading in dynamic enhanced clinical MR images, which provides a valuable reference for clinic diagnoses, treatments, and prognoses.

Conflict of interest

None.

Acknowledgments

This work was supported by the National Key Research and Development Plan (2017YFB1103602, 2018YFC0116904, 2017YFC0114304), National Natural Science Foundation of China (61801476), Scientific Research Equipment Program of Chinese Academy of Science (YJKYYQ20170050), Zhejiang Key Technology Research Development Program (2018C03024), Jiangsu Key Technology Research Development Program (BE2016613, BE2017675, BE2017663, BE2018610, BE2017664), Jiangsu Natural Science Foundation (BK20170387, BK20180221), Suzhou Industry Technological Innovation Projects (SYG201606, SYG201706, SYG201707), Suzhou Science & Technology Projects for People's Livelihood (SYS201656, SS201866, SYS2018010, SS201854, SS201855), Suzhou Science and Technology Development Project (SZS201609, SZS201818), SIBET Medical and Technology Project (Y853111305, Y853171305). We would also like to acknowledge the support provided the Central Hospital of Lishui, and The Second Affiliated Hospital of Soochow University for the clinical HCC dataset.

References

[1] S. Li, H. Jiang, W. Pang, Joint multiple fully connected convolutional neural network with extreme learning machine for hepatocellular carcinoma nuclei grading, Comput. Biol. Med. 84 (2017) 156–167.

[2] E. Goceri, Z.K. Shah, R. Layman, et al., Quantification of liver fat: a comprehensive review, Comput. Biol. Med. 71 (2016) 174–189.

[3] Y. Xie, J. Zhang, Y. Xia, et al., Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT, Inf. Fusion 42 (2018) 102–110.

[4] J. Sachdeva, V. Kumar, I. Gupta, et al., A package-SFERCB-"Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors", Appl. Soft Comput. 47 (2016) 151–167.

[5] E.I. Zacharaki, S. Wang, S. Chawla, et al., Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme, Magn. Reson. Med.: Off. J. Int. Soc. Magn. Reson. Med. 62 (6) (2009) 1609–1618.

[6] E. Malar, A. Kandaswamy, D. Chakravarthy, et al., A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine, Comput. Biol. Med. 42 (9) (2012) 898–905.

[7] S. Bonte, I. Goethals, R. Van Holen, Machine learning based brain tumour segmentation on limited data using local texture and abnormality, Comput. Biol. Med. 98 (2018) 39–47.

[8] P. Burlina, K.D. Pacheco, N. Joshi, et al., Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis, Comput. Biol. Med. 82 (2017) 80–86.

[9] Y. Xu, T. Mo, Q. Feng, et al., Deep learning of feature representation with multiple instance learning for medical image analysis, IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2014, pp. 1626–1630.

[10] A. Thiyagarajan, U.M. Pandurangan, Comparative analysis of classifier Performance on MR brain images, Int. Arab J. Inf. Technol. 12 (6A) (2015) 772–779.

[11] F.G. Zöllner, K.E. Emblem, L.R. Schad, SVM-based glioma grading: optimization by feature reduction analysis, Z. Med. Phys. 22 (3) (2012) 205–214.

[12] J. Sachdeva, V. Kumar, I. Gupta, et al., Multiclass brain tumor classification using GA-SVM, Developments in E-systems Engineering (DeSE), 2011, IEEE, 2011, pp. 182–187.

[13] F. Citak-Er, Z. Firat, I. Kovanlikaya, et al., Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T, Comput. Biol. Med. 99 (2018) 154–160.

[14] E. Goceri, N. Goceri, Deep learning in medical image analysis: recent advances and future trends, International Conference on Computer Science and Engineering (UBMK), vol 10, 2017, p. 5.

[15] Q. Wang, D. Que, Staging of Hepatocellular Carcinoma using deep feature in contrast-enhanced MR images, International Conference on Computer Engineering, Information Science & Application Technology, 2016.

[16] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[17] V. Bevilacqua, A. Brunetti, G.F. Trotta, et al., A deep learning approach for Hepatocellular Carcinoma grading, Int. J. Comput. Vis. Image Process. 7 (2) (2017) 1–18.

[18] K. Liu, G. Kang, Multiview convolutional neural networks for lung nodule classification, Int. J. Imag. Syst. Technol. 27 (1) (2017) 12–22.

[19] M.G. Ertosun, D.L. Rubin, Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks, AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2015, p. 1899.

[20] S. Khawaldeh, U. Pervaiz, A. Rafiq, et al., Noninvasive grading of Glioma tumor using magnetic resonance imaging with convolutional neural networks, Appl. Sci. 8 (1) (2017) 27.

[21] Y. Ishikawa, K. Washiya, K. Aoki, et al., Brain tumor classification of microscopy images using deep residual learning, SPIE BioPhotonics Australasia, International Society for Optics and Photonics, 201610013: 100132Y.

[22] K. Chang, H.X. Bai, H. Zhou, et al., Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR imaging, Clin. Canc. Res. 24 (5) (2018) 1073–1081.

[23] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[24] G. Huang, Z. Liu, L. Van Der Maaten, et al., Densely connected convolutional networks, CVPR, 1 2017, p. 3 (2).

[25] R. Dey, Z. Lu, Y. Hong, Diagnostic classification of lung nodules using 3D neural networks, Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. IEEE, 2018, pp. 774–778.

[26] Y. Chen, J. Li, H. Xiao, et al., Dual path networks, Advances in Neural Information Processing Systems, 2017, pp. 4467–4475.

[27] J. Hu, L. Shen, G. Sun, Squeeze-and-Excitation Networks, (2017), p. 7 1709.01507.

[28] H.A. Edmondson, P.E. Steiner, Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies, Cancer 7 (3) (1954) 462–503.

[29] L.H. Sobin, I.D. Fleming, TNM classification of malignant tumors, (1997), Cancer: Interdiscip. Int. J. Am. Canc. Soc. 80 (9) (1997) 1803–1804.

[30] M. Pirisi, M. Leutner, D.J. Pinato, et al., Reliability and reproducibility of the edmondson grading of hepatocellular carcinoma using paired core biopsy and surgical resection specimens, Arch. Pathol. Lab Med. 134 (12) (2010) 1818–1822.

[31] D. Gu, 3D densely Connected Convolutional Network for the Recognition of Human Shopping Actions, Université d'Ottawa/University of Ottawa, 2017.

[32] Y. Bengio, Practical Recommendations for Gradient-Based Training of Deep Architectures, Neural networks: Tricks of the trade Springer, Berlin, Heidelberg, 2012, pp. 437–478.

[33] L. Bottou, Stochastic Gradient Descent Tricks, Neural networks: Tricks of the trade Springer, Berlin, Heidelberg, 2012, pp. 421–436.