


RESEARCH ARTICLE

Open Access



# A hierarchical fusion strategy of deep learning networks for detection and segmentation of hepatocellular carcinoma from computed tomography images

I-Cheng Lee<sup>1,2†</sup>, Yung-Ping Tsai<sup>3†</sup>, Yen-Cheng Lin<sup>3</sup>, Ting-Chun Chen<sup>3</sup>, Chia-Heng Yen<sup>4</sup>, Nai-Chi Chiu<sup>5</sup>, Hsuen-En Hwang<sup>5</sup>, Chien-An Liu<sup>5</sup>, Jia-Guan Huang<sup>6</sup>, Rheun-Chuan Lee<sup>5</sup>, Yee Chao<sup>7</sup>, Shinn-Ying Ho<sup>3,8,9,10\*</sup> and Yi-Hsiang Huang<sup>1,2,11,12\*</sup> 

## Abstract

**Background** Automatic segmentation of hepatocellular carcinoma (HCC) on computed tomography (CT) scans is in urgent need to assist diagnosis and radiomics analysis. The aim of this study is to develop a deep learning based network to detect HCC from dynamic CT images.

**Methods** Dynamic CT images of 595 patients with HCC were used. Tumors in dynamic CT images were labeled by radiologists. Patients were randomly divided into training, validation and test sets in a ratio of 5:2:3, respectively. We developed a hierarchical fusion strategy of deep learning networks (HFS-Net). Global dice, sensitivity, precision and F1-score were used to measure performance of the HFS-Net model.

**Results** The 2D DenseU-Net using dynamic CT images was more effective for segmenting small tumors, whereas the 2D U-Net using portal venous phase images was more effective for segmenting large tumors. The HFS-Net model performed better, compared with the single-strategy deep learning models in segmenting small and large tumors. In the test set, the HFS-Net model achieved good performance in identifying HCC on dynamic CT images with global dice of 82.8%. The overall sensitivity, precision and F1-score were 84.3%, 75.5% and 79.6% per slice, respectively, and 92.2%, 93.2% and 92.7% per patient, respectively. The sensitivity in tumors < 2 cm, 2–3, 3–5 cm and > 5 cm were 72.7%, 92.9%, 94.2% and 100% per patient, respectively.

**Conclusions** The HFS-Net model achieved good performance in the detection and segmentation of HCC from dynamic CT images, which may support radiologic diagnosis and facilitate automatic radiomics analysis.

**Keywords** Hepatocellular carcinoma, Deep learning, Segmentation, Detection, Computed tomography

<sup>†</sup>I-Cheng Lee and Yung-Ping Tsai contributed equally to this work.

\*Correspondence:

Shinn-Ying Ho

syho@nycu.edu.tw

Yi-Hsiang Huang

yhhuang@vghtpe.gov.tw

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Hepatocellular carcinoma (HCC) is the sixth most common cancer and the fourth leading cause of cancer-related death in the world [1]. Dynamic computed tomography (CT) scan plays an important role in the diagnosis, staging and treatment decision of patients with HCC [2]. However, it requires professional and experienced radiologists to interpret hundreds slices of CT images per patient to detect HCC and determine the stage by measuring the size and number of HCC. After long working time, reading fatigue may lead to inefficiency of work, miss diagnosis, and even harm to physical and mental health [3]. Therefore, the artificial intelligence (AI)-assisted automatic detection and segmentation of HCC from dynamic CT images is in urgent need to assist radiologists. Besides, radiomics information from CT images may also have implication in predicting cancer outcomes [4]. Our recent study showed that by incorporating radiomic features in AI-derived prediction model could improve the prediction accuracy of HCC recurrence after resection [5]. The automatic segmentation of HCC may facilitate automatic radiomic analysis for AI-derived prediction models.

In the past, some studies have proposed automatic segmentation methods for liver tumors, but most of the studies used small datasets and only portal-venous phase CT images, which may not be sufficient to develop an accurate AI-detection model [6–13]. In clinical practice, radiologists read dynamic CT images including non-contrast, arterial, and portal-venous phase images slice-by-slice during interpretation. Due to the diverse heterogeneity of HCC, feature extraction methods and network architectures are important issues to such a difficult task. Model-based methods relied on handcrafted or non-handcrafted features of the image. With sufficient data and computational capability, deep learning may extract non-handcrafted features with better representation capability. Due to the powerful ability of deep learning, it could be easy to have overfitting by a single model. Using models of different learning strategies on a specific task in the method to collaborate might be a way to avoid the overfitting problem. The aim of this study is to develop a deep learning based network with multiple strategies for detection and segmentation of HCC from dynamic CT images.

## Materials and methods

### Patients

This study was approved by the Institutional Review Board, Taipei Veterans General Hospital, which complied with standards of the Declaration of Helsinki and current ethical guidelines. Due to the retrospective nature of the study, the Institutional Review Board waived the need for written informed consent. The identifying information

of the enrolled subjects has been delinked and therefore authors could not access the information.

From October 1, 2007 to August 31, 2019, 884 consecutive HCC patients receiving surgical resection and radiofrequency ablation (RFA) in Taipei Veterans General Hospital with available CT images before treatment were retrospectively screened. The inclusion criteria were: (1) Age  $\geq 20$  years; (2) Available CT image within 3 months prior to surgical resection or RFA; (3) No other loco-regional treatment prior to surgical resection or RFA. Patients were excluded by the following criteria: (1) without complete dynamic contrast-enhanced CT (CECT) images including non-contrast phase, arterial phase and portal venous phase ( $n=87$ ); (2) poor image quality or unable to align the dynamic CT images according to the Z-axis ( $n=202$ ). Finally, 595 HCC patients with complete dynamic CECT images were enrolled in this study. HCC was diagnosed before surgery or RFA by CECT or magnetic resonance imaging (MRI), which fulfilled the diagnostic criteria of the American Association for the Study of Liver Diseases (AASLD) treatment guidelines for HCC [14] or was confirmed pathologically after surgery and RFA.

### CECT image segmentation

The image acquisition protocols of CT scanners involved in the present study are shown in Table S1. Interpretation and tumor segmentation of all CECT images of the 595 patients were performed by three radiologists who were blinded to the clinical and pathological data. The labeling of the HCC in dynamic CT images included the non-contrast phase, arterial phase and portal-venous phase. The three radiologists had read  $>2,000$  liver CT studies per year for at least 5 years. When contouring the tumor, the edge of the observed focal lesion within the liver was defined as an imaging appearance that is distinctive from the background according to the Liver Reporting and Data System (LI-RADS) [2, 15]. The contours of the liver were also labelled in 200 cases for training of segmentation of the liver. For evaluation of the HFS-Net model, ground truth tumor compartments were delineated manually. This was performed using a semiautomatic approach with subsequent manual editing (IntelliSpace Discovery; Philips Healthcare, Netherlands), performed by the three experienced radiologists [5].

### CT image processing

The characteristics of the CT image dataset were shown in Table 1. The original CT images were all 3D images stacked by  $512 \times 512$  2D slices. Before the experiment, this study used downsampling to reduce all slices into a size of  $256 \times 256$  and adjusted all CT image scale to 1 pixel equal to 1.4 mm. We aligned the dynamic CT images according to the Z-axis coordinates of the slices

**Table 1** Characteristics of the hepatocellular carcinoma

Dataset	Case number	Tumor size (cm)				Tumor number		BCLC stage		
		< 2	2–3	> 3–5	> 5	1	> 1	0	A	B
All	595	93	133	178	191	500	95	83	440	72
Training	298	42	78	81	97	245	53	38	220	40
Validation	118	18	27	45	28	104	14	14	95	9
Test	179	33	28	52	66	151	28	31	125	23

and superimposed them into a three-channel image for use. After alignment, 24,810 images (a total of 74,430 slices) were enrolled for analysis.

### Study design for HFS-Net

To design the best architecture of HFS-Net for HCC detection and segmentation, an ablation study was first conducted for evaluating various architectures of deep learning models. We randomly selected 491 cases from a total of 595 patients with a broad distribution of various tumor sizes was conducted for the ablation study.

The whole dataset was used to design and evaluate HFS-Net, which was randomly divided into non-overlapping training, validation, and test datasets in a 5:2:3 ratio. The training, validation, and test sets have 298, 118, and 179 patients, respectively. Each patient had a different number of slices and all were analyzed. The training set is used for constructing the HFS-Net model, while the validation set is used for tuning and optimization of model parameters. We then evaluate the performance of the HFS-Net model on a separate test set. HFS-Net was expected to combine the best sub-models by thorough evaluation of the candidate sub-models.

Therefore, the splitting strategy of the data and the construction of the model are independent in the two stages, and the purposes in the two stages are different. There is no comparison between the experimental results from the two stages, and the constructed models in the ablation experiments are not used in the final HFS-Net model. Instead, it uses the conclusions drawn from the ablation experiments as the background knowledge for model construction.

### Ablation study design for HFS-Net development

The ablation study investigates the performance of the proposed AI system HFS-Net by evaluating certain sub-models to understand the contribution and tolerance of the sub-models to the overall system. This study aims to discern the individual and combined strength of various sub-models in various scenarios. The influencing factors of system performance include architectures (U-Net, DenseU-net, Hyper-DenseU-Net, and R2U-Net), phases (non-contrast, arterial, portal-venous, and dynamic phases), loss functions (cross entropy, focal loss, median frequency balance, and dice loss), tumor sizes (the longest tumor axes in slices), assessment (dice score for

model's segmentation ability and tumor detection rate) and the fusion strategy.

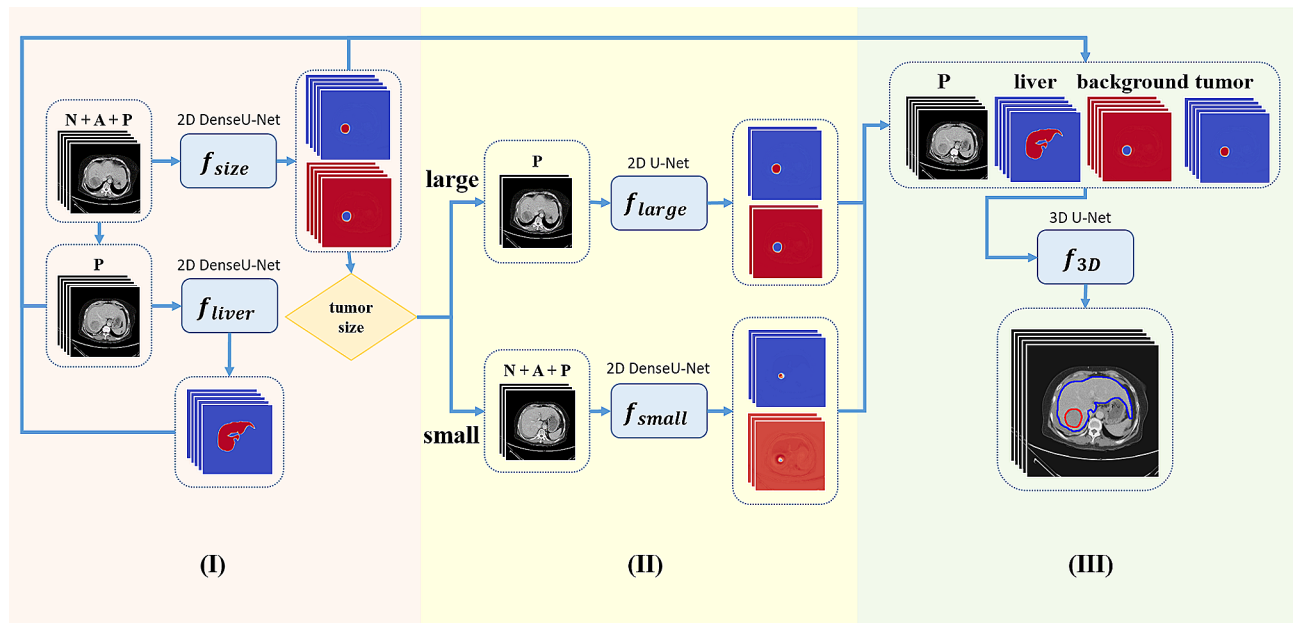
The final stage of the HFS-Net design involves a comparison between 2D HFS-Net and the more comprehensive 3D HFS-Net. 2D HFS-Net leveraged the initial stages of hierarchical architecture and 2-D spatial features, while 3D HFS-Net incorporates a complete hierarchical architecture with fusion strategy and 3-D spatial features. This comparison helps to thoroughly evaluate the impact of hierarchical structures and fusion strategies to enhance the segmentation and detection of HCC.

### Neural network architecture of HFS-Net

Figure 1 showed the data flow of our proposed HFS-Net method for liver and tumor detection and segmentation according to results of the ablation study. We cascaded five sub-models trained by different learning strategies (Table 2). HFS-Net consisted of three stages based on U-Net and DenseU-Net. The first stage of HFS-Net is the liver and tumor segmentation with tumor size estimation using 2D DenseU-Net, where every slice of the entire CT scan case was taken as input. The second stage is the divide-and-conquer stage. According to the tumor size calculated in the first stage, the tumors with the longest axis less than or equal to  $m$  pixels in the slice are assigned to the smaller tumor group, and those with more than  $m$  pixels are assigned to the larger tumor group. This study uses  $m=30$  pixels as the demarcation point, which is about 4.2 cm. In the divide-and-conquer stage, we use a customized model to adaptively segment large and small tumors in slices. For small-tumor groups, 2D DenseU-Net uses dynamic CT images to segment tumors, and for large-tumor groups, 2D U-Net uses portal-venous phase CT images to segment tumors. The third stage is the fusion strategy stage. This stage integrates the portal-venous phase CT image, the outcomes of the previous two stages and the segmentation of the liver, and uses 3D U-Net to segment the final result of the 3D liver tumor. The detailed modeling and learning strategies of HFS-Net are described in Supplementary Methods.

### Evaluation of the performance of the HFS-Net model

We used dice per case and dice global of Jaccard similarity as metrics to evaluate the performance of segmentation. Dice per case represents the average dice of the case, and dice global is the dice score that combines all the



**Fig. 1** The HFS-Net data flow for liver and tumor segmentation. Stage I: Identify tumor's longest axis in every slice of a case. Stage II: Accord tumor size by feeding the CT slice which the longest axis of tumors over 30 pixels in the slice for  $f_{large}$  computation and which the longest axis less than 30 pixels of tumors in the slice for  $f_{small}$  computation. Stage III: Combine venous phases of CT images and results of  $f_{liver}$ ,  $f_{size}$ ,  $f_{large}$  and  $f_{small}$  as input of  $f_{3D}$  computation for getting final segmentation of tumors

**Table 2** Function strategy in the HFS-Net

	Phase	Neural network	Loss function
$f_{liver}$	Portal-venous	2D DenseU-Net	cross entropy + dice loss
$f_{size}$	Dynamic	2D DenseU-Net	focal loss + dice loss
$f_{large}$	Portal-venous	2D U-Net	cross entropy
$f_{small}$	Dynamic	2D DenseU-Net	focal loss + dice loss
$f_{3D}$	Portal-venous	3D U-Net	focal loss + dice loss

Dynamic CT images include non-contrast, arterial, and portal-venous phases

slices as a case to calculate. We used sensitivity, precision, and F1-score to evaluate detection performance. When evaluating the performance of detection, the criterion for successful detection is defined as the dice global that overlaps the model results with the corresponding tumor labels exceeding  $\theta$ . Otherwise, it is a false positive. If the model fails to detect a tumor in a slice, it is considered a false negative. We set  $\theta=0.2$  to require a significant, but not exact overlap as post study did [16]. In order to analyze the detection effect of the model more deeply, we designed four indicators to evaluate the detection performance of the model, including overall detection performance of tumors in slices (Per tumor volume), mean detection performance of tumors in slices (Per tumor cut), abnormal slice detection (Per slice) and abnormal case detection (Per patient).

The overall detection performance of tumors in slices means to use the tumors in all slices as the denominator to evaluate the performance of tumor detection. The mean detection performance of tumors in slices means to use the tumors in all slices of the case as the denominator

to evaluate the performance of each case on average. Abnormal slice detection means the ability to detect at least one of the tumors in a slice with tumors, and abnormal case detection means the ability to detect at least one of the tumors in a case with tumors.

### Experiment environment

In this study, we used Nvidia GeForce RTX 2080 Ti (12GB) as the GPU, Intel(R) Xeon(R) Gold 6136 CPU @ 3.00 Ghz as the CPU, and used the CentOS Linux release 8.3.2011 as operating system. The method was implemented with python3.7 and Pytorch packages [17].

## Results

### HCC characteristics

The characteristics of the HCC were shown in Table 1. Among the 595 patients with HCC, 93 (15.6%), 133 (22.4%), 178 (29.9%) and 191 (32.1%) cases had HCC smaller than 2 cm, 2 to 3 cm, larger than 3 to 5 cm, and larger than 5 cm, respectively. Majority of the HCC were single tumor (84%) and in the BCLC stages 0 and A (13.9% and 73.9%, respectively).

### HFS-Net model development and fusion strategy approach

To evaluate the performance of the customized sub-models of HFS-Net in specific tumor size, we used 294 cases (37,608 slices) of CT images to train the model, 61 cases (7,833 slices) for validation, and 136 cases (12,576 slices) for test in the first part of the experiment. The longest axis of the tumor in the slice was used as the size basis. At

first, we developed a 2D DenseU-Net model for liver segmentation, which may facilitate further HCC detection within the liver. The dice per case and dice global of the 2D DenseU-Net model for liver segmentation were 95.2% and 95.3%, respectively, in the test set (Table S2).

We applied U-Net and DenseU-Net to CT images of various phases to explore the phase effects on the deep learning models. The U-Net model using the portal-venous phase images performed best for tumors above 30 pixels, whereas the DenseU-Net model using dynamic CT images performed best for tumors with a size of 10 to 30 pixels (Table S3). The DenseU-Net model with dynamic CT images achieved the highest dice score of 48.9%. The second-highest performance was achieved by the DenseU-Net model using portal-venous phase images, with a dice score of 43.83%.

Then we used the same dataset to test the performance of models using different neural network architectures, including U-Net, DenseU-Net, R2U-Net, Hyper-DenseU-Net, and Single Dense Path U-Net. The latter two are designed for multimodal medical imaging [18, 19]. The results showed that the U-Net model using portal-venous phase images performed best for tumors above 30 pixels, achieving an average dice score of 87.56% in large tumor segmentation, whereas the DenseU-Net model using dynamic CT images performed best for tumors with a size of 10 to 30 pixels (Table S4).

In the next step, we tested the U-Net and DenseU-Net models with different loss functions, including cross-entropy, focal loss, median frequency balance focal loss, and dice loss. The results showed that the DenseU-Net model using dynamic CT images with focal loss plus dice loss as the loss function could improve the segmentation performance for smaller tumors, while the U-Net model using portal-venous phase images with cross-entropy performed best for larger tumors (Table S5).

Finally, in order to test whether the hierarchical fusion strategy of HFS-Net is effective, we evaluated the test performance of 2D HFS-Net and 3D HFS-Net (Table S6). The 2D HFS-Net means that HFS-Net contains hierarchical but does not contain fusion strategy, whereas the 3D HFS-Net used both hierarchical and fusion strategies. The results showed that the 2D HFS-Net model could

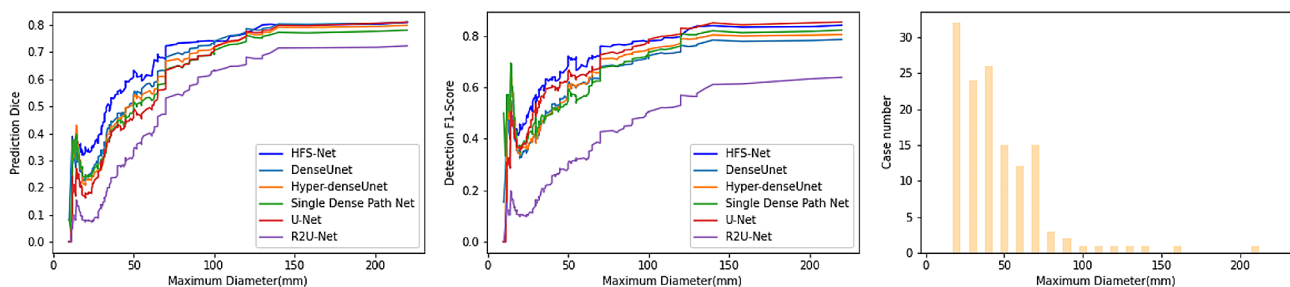
improve the segmentation performance and detection rate of tumors with a size of 10 to 50 pixels after combining U-Net and DenseU-Net models. Since the 3D HFS-Net model considered the features of the 3D space and integrated the outcomes of models of different strategies, the 3D HFS-Net model not only improved the segmentation performance and detection rate as compared to the 2D HFS-Net model, but also greatly reduced the false detection rate. The data flow of the finally proposed HFS-Net model for liver and tumor detection and segmentation was shown in Fig. 1.

### Performance of the HFS-Net model

In this section, we randomly divide the complete dataset in a ratio of 5:2:3, used 298 cases (37,824 slices) of CT images to train the model, 118 cases (15,423 slices) for validation, and 179 cases (22,803 slices) for test. The training set fits the model, while the validation set tunes and optimizes the model. The final HFS-Net model was developed from the training and validation sets. We subsequently evaluate the performance of the model on a separate test set. The performance of the HFS-Net model for HCC detection and segmentation in the training, validation and test sets was shown in Table S7.

We compared the HFS-Net model with other single-strategy models by using the test set. As shown in Fig. 2, the HFS-Net model generally had the highest prediction dice and F1-score compared to other single-strategy models, suggesting that the HFS-Net model outperforms other single-strategy models for the detection and segmentation of HCC from small to large tumor size.

The performance of the HFS-Net model in the test set was shown in Table 3. The overall dice per case was 58.7%, and the dice global was 82.8%. The overall sensitivity, precision and F1-score were 84.3%, 75.5% and 79.6% per slice, respectively, and were 92.2%, 93.2% and 92.7% per patient, respectively. Notably, in analysis per patient, there is no negative case in the prediction. Therefore, the representativeness of the F1-score would be similar to that of the sensitivity. The sensitivity, precision and F1-score per patient of the HFS-Net model were generally higher than 90% for tumors larger than 2 cm. For tumors less than 2 cm, the accuracy of the HFS-Net model



**Fig. 2** Performance of HFS-Net segmentation (left), detection (middle) and distribution of sizes of HCC (right)

**Table 3** Performance of HFS-Net in the test set

		Segmentation		Per tumor volume	Per tumor cut	Per slice	Per patient
All ( <i>n</i> = 179)	Dice per case	58.7%	Sensitivity	88.9%	82.1%	84.3%	92.2%
	Dice global	82.8%	Precision	51.5%	67.6%	75.5%	93.2%
	MTD MAE	0.73 cm	F1-Score	65.3%	74.2%	79.6%	92.7%
1 ~ 2 cm ( <i>n</i> = 33)	Dice per case	32.0%	Sensitivity	69.4%	52.4%	57.0%	72.7%
	Dice global	32.7%	Precision	45.5%	34.1%	39.5%	77.4%
	MTD MAE	0.56 cm	F1-Score	54.9%	41.3%	46.7%	75.0%
2 ~ 3 cm ( <i>n</i> = 28)	Dice per case	47.3%	Sensitivity	90.3%	75.0%	77.8%	92.9%
	Dice global	55.5%	Precision	57.1%	58.7%	66.7%	92.9%
	MTD MAE	0.57 cm	F1-Score	70.0%	65.9%	71.7%	92.9%
3 ~ 5 cm ( <i>n</i> = 52)	Dice per case	60.4%	Sensitivity	89.1%	75.7%	79.8%	94.2%
	Dice global	69.1%	Precision	49.0%	60.8%	69.0%	94.2%
	MTD MAE	0.48 cm	F1-Score	63.2%	67.4%	74.0%	94.2%
> 5 cm ( <i>n</i> = 66)	Dice per case	76.1%	Sensitivity	98.5%	88.1%	89.5%	100%
	Dice global	84.5%	Precision	54.0%	75.6%	84.2%	100%
	MTD MAE	1.04 cm	F1-Score	65.7%	81.4%	86.8%	100%
1 tumor ( <i>n</i> = 151)	Dice per case	58.2%	Sensitivity	89.2%	81.7%	83.7%	92.7%
	Dice global	84.2%	Precision	52.0%	66.7%	74.7%	93.3%
	MTD MAE	0.70 cm	F1-Score	65.7%	73.4%	78.9%	93.0%
> 1 tumors ( <i>n</i> = 28)	Dice per case	62.9%	Sensitivity	87.5%	83.3%	86.0%	89.3%
	Dice global	78.1%	Precision	49.1%	70.6%	78.2%	92.6%
	MTD MAE	0.90 cm	F1-Score	62.9%	76.4%	82.0%	90.9%

MTD, maximum tumor diameter; MAE, mean absolute error

decreases but the sensitivity and precision per patient remained higher than 70%. Figure 3 showed the examples of HCC segmentation by the HFS-Net model in the test set. Our model could detect majority of the tumors with various sizes, although there were some differences in the tumor margin between the labelling by radiologists and the HFS-Net model (Fig. 3A-C). Some tumors could be detected in the slices of maximum tumor size but might be missed in the marginal slice of the tumor (Fig. 3D). The HFS-Net model could also detect few small tumors in individual CT slices which were initially missed by the radiologist's labeling (Fig. 3E), and the labeling of these missed tumors could be corrected.

#### Calculating time

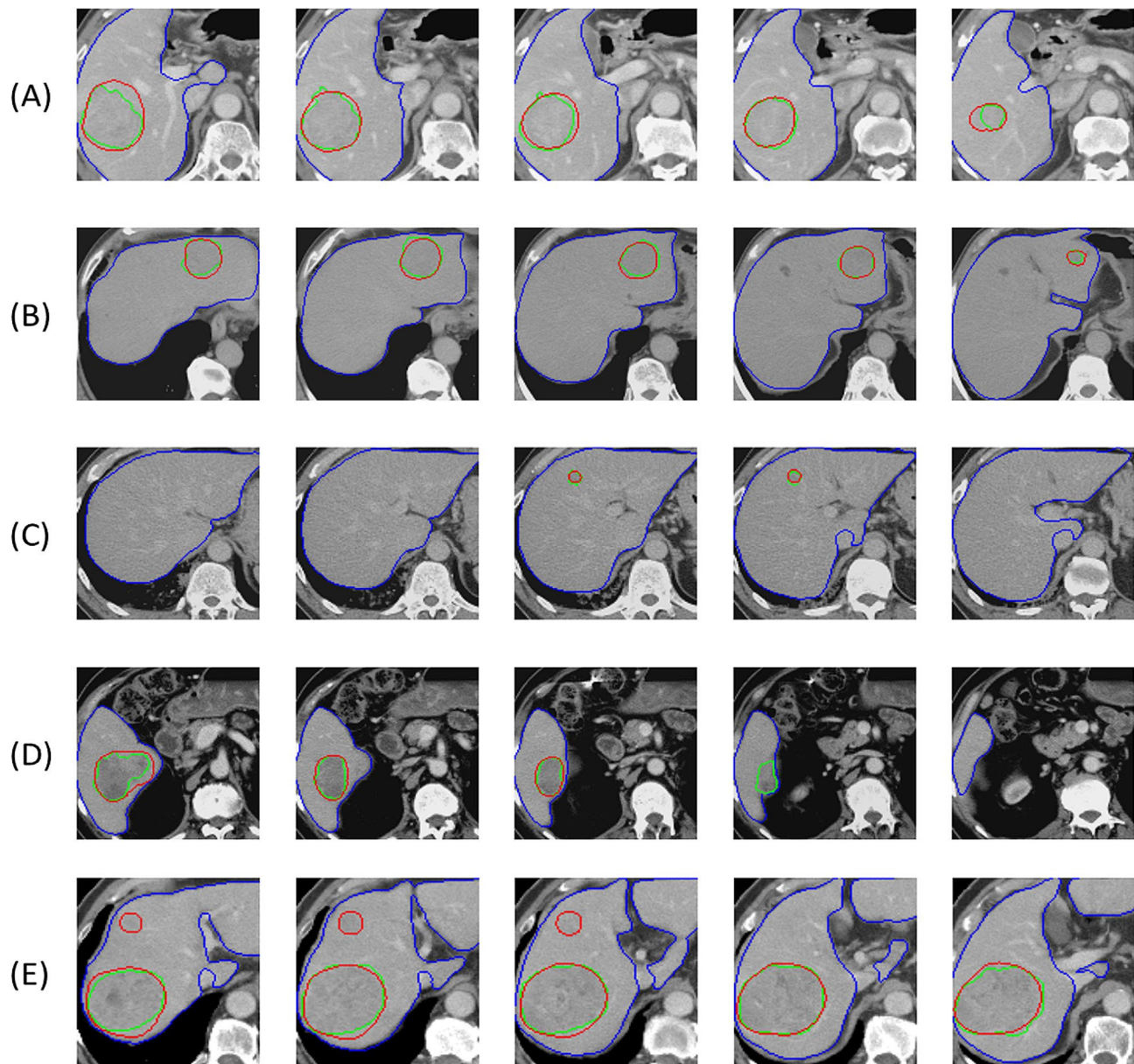
HFS-Net inference was mainly divided into three stages: (1) reading model parameters and moving to GPU memory; (2) reading CT images to be tested; (3) the model inference. It took about 1.527 s to read the parameters trained by HFS-Net to the CPU, and about 3.453 s to copy HFS-Net from the CPU to the GPU. It took about 0.028 s to read the CT image to be tested, convert it into an array from dicom format and perform pre-processing for each slice. It took about 0.030 s to copy the array of the image to be tested to the GPU and perform inference for each slice. The step of reading model parameters only needs to be read once during the initial state, and multiple CT image inferences can be performed with once reading. Across the entire dataset, the average time for reading data to complete inference was 4.517 s, with a

standard deviation of 0.877 s. By testing a case with 126 slices, the real execution time was 4.29 s. The demonstration of the HFS-Net model is shown in the supplementary movie. From the data presented, it is clear that the HFS-Net can complete its processing within 10 s.

#### Discussion

In this study, we proposed a hierarchical fusion strategy of deep learning network called HFS-Net, which used a coarse-to-fine mechanism to automatically detect HCC from CT images. We found that dynamic CT images combined with deep learning methods could perform better than a single-phase image model in detecting and segmenting small tumors. We used the hierarchical fusion strategy to solve the single-model trade-off between small and large tumor segmentation of CT slices. Compared with the single-strategy model, our method has improved detection and segmentation performance in tumors of various sizes. Our method achieves acceptable performance for detecting and segmenting HCC and can interpret a patient's CT scans within 10 s.

This study explored the effect of the dynamic CT model and proposed a novel deep learning architecture to segment liver and HCC in CT scans. HCC may show different brightness in different phases of CT images. Interpreting CT images of non-contrast, arterial and portal-venous phases is necessary for the clinical judgment of liver tumors. Although deep learning methods have achieved good results in the field of tumor segmentation



**Fig. 3** Examples of HFS-Net segmentation results from test dataset. **A** A 8.5 cm HCC representing large HCC (dice per case 88.1%). **B** A 4.6 cm HCC representing medium sized HCC (dice per case 80%). **A** A 8.5 cm HCC representing large HCC (dice per case 88.1%). **C** A 1.7 cm HCC representing BCLC stage 0 HCC (dice per case 83.6%). **D** A tumor was detected in the first three slices but missed in the marginal slice of the tumor. **E** A large tumor accurately segmented by HFS-Net, and a small tumor missed by radiologist labelling but detected by HFS-Net. The green line represents the radiologist's label, the red line represents the output of HFS-Net for tumor margin, and the blue line represents the output of HFS-Net for liver margin

in the past, they lacked consideration of different phase features of CT images [6–13]. In this study, we found that using dynamic CT images with DenseU-Net could improve the performance in segmenting smaller tumors. DenseU-Net's dense block uses tightly connected paths to achieve an effective dynamic CT feature extraction method. A previous study indicated that DenseU-Net combined with focal loss is quite effective as a model architecture for small target segmentation [20].

U-Net is a popular method for image segmentation and has achieved success in many medical image segmentation tasks [21–23]. However, U-Net, which has less interaction of high-level features and low-level features, could not effectively learn the features of dynamic CT images. We found that U-Net using single portal-venous phase CT images to segment larger HCC had better performance. HCC in the portal-venous phase had clearer edges and were less likely to be confused by noise when segmenting larger tumors. In order to maximize

the performance for both large and small tumors, we proposed a novel hierarchical fusion strategy of deep learning networks, which uses multi-task learning to customize suitable learning strategies for specific tasks and merge these strategies at the end to output more reliable results. Our results showed that the model segmentation and detection performance in each tumor size is better than other models using a single learning strategy.

The HFS-Net model could achieve a good performance in detecting HCC, and the sensitivity, precision and F1-score per patient were generally higher than 92%. When stratified by tumor size, the accuracy for detecting HCC larger than 2 cm remained higher than 92%, whereas the accuracy of the HFS-Net model decreases for tumors less than 2 cm, with the sensitivity and precision per patient for HCC less than 2 cm of 72.7% and 77.4%, respectively. Very early HCC may have less obvious enhancing pattern and may be difficult to be diagnosed by CT scan [24]. In clinical practice, some of the atypical HCC presented in CT scan were alternatively diagnosed by MRI or confirmed pathologically after resection or RFA [25]. More case numbers and additional image modalities might be needed to further improve the detection accuracy for small HCC. The imperfect detection rates of the HFS-Net model, especially for HCC less than 2 cm, indicate that the current AI-derived model could assist but not replace the interpretation of the radiologists.

The dice global of the HFS-Net model was 82.8%, indicating an acceptable accuracy of tumor segmentation. An accurate tumor segmentation may assist automatic measurement of tumor size and volume, which are important clinical parameters for staging, treatment decision and outcomes prediction [26, 27]. Moreover, automatic segmentation may facilitate radiomics analysis of HCC. Our recent study showed that the radiomics of HCC may contain significant prognostic information and radiomics analysis is increasingly adopted by AI-derived prognostic models for cancer outcomes [4, 5]. However, manual labeling of the liver and HCC were quite exhausting and time-consuming. When establishing the HCC labeling dataset in this study, it took about 1 min to interpret liver CT for detection of HCC and 10 min for labeling the tumor in multiple slices by a radiologist. After development of the HFS-Net model, it took only about 8 s for automatic detection and segmentation of HCC for one patient.

This study has some limitations. First, this is a retrospective study from a single center. The accuracy of the HFS-Net model need further external validation in the future. Second, this study only analyzed CT images with HCC. The detection ability for other benign and malignant liver tumors was not evaluated in this study. Clinical image datasets of different liver tumors to train the

AI model for detection and differential diagnosis of other types of liver tumors are needed in the future. Furthermore, while the handling of breathing artifacts is an issue, our methodology did not explicitly focus on correcting or compensating for these artifacts. Therefore, incorporating image registration into the enhancement of the image processing methods may advance model performance. Third, although the dice global of the HFS-Net model achieved 82.8%, the accuracy was relatively lower for tumors with smaller size. Further research is needed to improve prediction accuracy for smaller tumors. Fourth, this study only included CT images from patients with earlier stage HCC prior to resection or RFA. Therefore, the current model could only detect intrahepatic HCC before treatment. Future works are needed to analyzed post-treatment CT images as well as extrahepatic lesions and vascular invasion.

In conclusion, by establishing a large dataset of CT images with HCC labeling, we developed a novel HFS-Net model, which performed better than other single-strategy models for both small and large HCC. The HFS-Net model could achieve a good performance in automatic detection and segmentation of HCC from dynamic CT images in only a few seconds. This model may support radiologic diagnosis and facilitate automatic radiomics analysis in the future.

#### Abbreviations

AASLD	American Association for the Study of Liver Diseases
AI	artificial intelligence
CT	computed tomography
CECT	contrast-enhanced computed tomography
HCC	hepatocellular carcinoma
HFS-Net	hierarchical fusion strategy of deep learning networks
RFA	radiofrequency ablation

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40644-024-00686-8>.

Supplementary Material 1

Supplementary Material 2

#### Acknowledgements

The authors thank the Clinical Research Core Laboratory, Taipei Veterans General Hospital for providing their facilities to conduct this study.

#### Funding

The study was supported by grants from Taipei Veterans General Hospital, Taipei, Taiwan (V109E-008-2, V109E-008-2(110), V111C-114, V112C-126), National Science and Technology Council, Taiwan (MOST 108-3011-F-075-001-110-2221-E-A49-099-MY3, 111-2740-B-400-002-, 109-2314-B-010-034-MY3, 110-2314-B-075-052, 111-2314-B-075-055, NSTC 112-2628-B-A49-016-MY3), and Veterans General Hospitals and University System of Taiwan Joint Research Program (VGHUST113-G1-9-2), and was financially supported by the "Center for Intelligent Drug Systems and Smart Bio-devices (IDS<sup>2</sup>B)" from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.



### Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Declarations

#### Ethics approval

This study was approved by the Institutional Review Board, Taipei Veterans General Hospital (IRB number: 2022-08-015AC), which complied with standards of the World Medical Association Declaration of Helsinki and current ethical guidelines.

#### Consent to participate

Due to the retrospective nature of the study, the Institutional Review Board waived the need for written informed consent. The identifying information of the enrolled subjects has been delinked and therefore authors could not access the information.

#### Consent for publication

Not applicable.

#### Competing interests

YHH has received research grants from Gilead Sciences and Bristol-Meyers Squibb, and honoraria from Abbvie, Gilead Sciences, Bristol-Meyers Squibb, Ono Pharmaceutical, Merck Sharp & Dohme, Eisai, Eli Lilly, Ipsen, and Roche, and has served in an advisory role for Abbvie, Gilead Sciences, Bristol-Meyers Squibb, Ono Pharmaceuticals, Eisai, Eli Lilly, Ipsen, Merck Sharp & Dohme and Roche. The other authors declare no conflicts of interest.

#### Author contribution statement

Lee IC, Tsai YP: data acquisition; analysis and interpretation of data; drafting of the manuscript.

Tsai YP, Chen TC, Lin YC, Yen CH: implementation and analysis of all deep learning methods.

Chiu NC, Liu CA, Hwang HN, Huang JG, Chau GY, Lee RC, Chao Y: data acquisition.

Ho SY, Huang YH: study concept and design; obtained funding; critical revision of the manuscript for important intellectual content; study supervision.

#### Author details

<sup>1</sup>Division of Gastroenterology and Hepatology, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>2</sup>School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

<sup>3</sup>Institute of Bioinformatics and Systems Biology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

<sup>4</sup>Institute of Computer Science and Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

<sup>5</sup>Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>6</sup>National Taiwan University School of Medicine, Taipei, Taiwan

<sup>7</sup>Cancer Center, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>8</sup>Department of Biological Science and Technology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

<sup>9</sup>Center for Intelligent Drug Systems and Smart Bio-devices (IDS 2 B), National Yang Ming Chiao Tung University, Hsinchu, Taiwan

<sup>10</sup>College of Health Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan

<sup>11</sup>Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

<sup>12</sup>Healthcare and Service Center, Taipei Veterans General Hospital, Taipei, Taiwan

Received: 13 May 2023 / Accepted: 8 March 2024

Published online: 26 March 2024

### References

1. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers*. 2021;7(1):6.

- Moura Cunha G, Chernyak V, Fowler KJ, Sirlin CB. Up-to-date role of CT/MRI LI-RADS in Hepatocellular Carcinoma. *J Hepatocell Carcinoma*. 2021;8:513–27.
- Waite S, Kolla S, Jeudy J, Legasto A, Macknik SL, Martinez-Conde S, et al. Tired in the reading room: the influence of fatigue in radiology. *J Am Coll Radiol*. 2017;14(2):191–7.
- Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol*. 2022;19(2):132–46.
- Lee IC, Huang JY, Chen TC, Yen CH, Chiu NC, Hwang HE, et al. Evolutionary learning-derived clinical-Radiomic models for Predicting Early recurrence of Hepatocellular Carcinoma after Resection. *Liver Cancer*. 2021;10(6):572–82.
- Li X, Chen H, Qi X, Dou Q, Fu C-W, Heng P-A. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging*. 2018:2663–74.
- Wardhana G, Naghibi H, Sirmacek B, Abayazid M. Toward reliable automatic liver and tumor segmentation using convolutional neural network based on 2.5 D models. *Int J Comput Assist Radiol Surg*. 2021:41–51.
- Chen L, Song H, Wang C, Cui Y, Yang J, Hu X et al. Liver tumor segmentation in CT volumes using an adversarial densely connected network. *BMC Bioinformatics*. 2019:1–13.
- Chlebus G, Schenk A, Moltz JH, van Hahn GB, Meine HK. H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci Rep*. 2018:1–7.
- Budak Ü, Guo Y, Tanyildizi E, Şengür A. Cascaded deep convolutional encoder-decoder neural networks for efficient liver tumor segmentation. *Med Hypotheses*. 2020:109431.
- Meng L, Tian Y, Bu S. Liver tumor segmentation based on 3D convolutional neural network with dual scale. *J Appl Clin Med Phys*. 2020;144:57.
- Jiang H, Li S, Li S. Registration-based organ positioning and joint segmentation method for liver and tumor segmentation. *BioMed research international*. 2018.
- Anter AM, Hassenian AE. CT liver tumor segmentation hybrid approach using neutrosophic sets, fast fuzzy c-means and adaptive watershed algorithm. *Artif Intell Med*. 2019:105–17.
- Heimbach JK, Kulik LM, Finn RS, Sirlin CB, Abecassis MM, Roberts LR, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67(1):358–80.
- Spieler B, Sabottke C, Moawad AW, Gabr AM, Bashir MR, Do RKG et al. Artificial intelligence in assessment of hepatocellular carcinoma treatment response. *Abdom Radiol (NY)*. 2021.
- Chlebus G, Schenk A, Moltz JH, van Ginneken B, Hahn HK, Meine H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci Rep*. 2018;8(1):1–7.
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32:8026–37.
- Alom MZ, Yakopcic C, Taha TM, Asari VK, editors. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). *NAECON 2018-IEEE National Aerospace and Electronics Conference*; 2018: IEEE.
- Dolz J, Gopinath K, Yuan J, Lombaert H, Desrosiers C, Ayed IB. HyperDenseNet: a hyper-densely connected CNN for multi-modal image segmentation. *IEEE Trans Med Imaging*. 2018;38(5):1116–26.
- Dong R, Pan X, Li F. DenseU-net-based semantic segmentation of small objects in urban remote sensing images. *IEEE Access*. 2019;7:65347–56.
- Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat Methods*. 2019;16(1):67–70.
- Ronneberger O, Fischer P, Brox T, editors. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*; 2015: Springer.
- Sousa JA, Paiva A, Silva A, Almeida JD, Braz Junior G, Diniz JO, et al. Automatic segmentation of retinal layers in OCT images with intermediate age-related macular degeneration using U-Net and DexiNed. *PLoS ONE*. 2021;16(5):e0251591.
- Roberts LR, Sirlin CB, Zaiem F, Almasri J, Prokop LJ, Heimbach JK, et al. Imaging for the diagnosis of hepatocellular carcinoma: a systematic review and meta-analysis. *Hepatology*. 2018;67(1):401–21.
- Vilgrain V, Van Beers BE, Pastor CM. Insights into the diagnosis of hepatocellular carcinomas with hepatobiliary MRI. *J Hepatol*. 2016;64(3):708–16.

26. Lee IC, Lei HJ, Chau GY, Yeh YC, Wu CJ, Su CW, et al. Predictors of long-term recurrence and survival after resection of HBV-related hepatocellular carcinoma: the role of HBsAg. *Am J Cancer Res.* 2021;11(7):3711–25.
27. Hung YW, Lee IC, Chi CT, Lee RC, Liu CA, Chiu NC, et al. Redefining Tumor Burden in patients with Intermediate-Stage Hepatocellular Carcinoma: the Seven-Eleven Criteria. *Liver Cancer.* 2021;10(6):629–40.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.