A New Three-stage Curriculum Learning Approach for Deep Network Based Liver Tumor Segmentation

1st Huiyu Li
School of Computer Science
Beijing Institute of Technology
Beijing, China
lihuiyu@bit.edu.cn

2nd Xiabi Liu
School of Computer Science
Beijing Institute of Technology
Beijing, China
liuxiabi@bit.edu.cn

3rd Said Boumaraf
School of Computer Science
Beijing Institute of Technology
Beijing, China
said.boumaraf@bit.edu.cn

4th Weihua Liu
School of Computer Science
Beijing Institute of Technology
Beijing, China
liuweihua@bit.edu.cn

5th Xiaopeng Gong
School of Computer Science
Beijing Institute of Technology
Beijing, China
gongxp@bit.edu.cn

6th Xiaohong Ma
Department of Diagnostic Radiology
Cancer Hospital, Chinese Academy of Medical Sciences
Beijing, China
maxiaohong@cicams.ac.cn

***Abstract*—Automatic segmentation of liver tumors in medical images is crucial for computer-aided diagnosis and therapy. It is a challenging task, since the tumors are notoriously small against the background voxels. This paper proposes a new three-stage curriculum learning approach for training deep networks to tackle this small object segmentation problem. The learning in the first stage is performed on the whole input volume to obtain an initial deep network for tumor segmentation. Then the second stage of learning focuses on the tumor-specific features by continuing training the network on the tumor patches. Finally, we retrain the network on the whole input volume in the third stage, in order that the tumor-specific features and the global context can be integrated to improve the final segmentation accuracy. With this approach, we can employ a single network to segment the tumors directly without the need of liver segmentation. We evaluate our approach on a clinical dataset from the hospital and the public MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge dataset. In the experiments, our approach exhibits significant improvement compared with the commonly used cascade counterpart.**

***Index Terms*—Liver Tumor Segmentation, CT, Curriculum Learning, Deep Learning**

I. Introduction

Liver cancer is the second most common cause of cancer death worldwide. Computed Tomography (CT) is the preferred imaging modality for tumor diagnosis and treatment. In clinical practices, segmenting malignant tissues is a prerequisite step for final cancer diagnosis and treatment planning. However, manual segmentation is time-consuming and poorly reproducible. An accurate and automatic method of liver tumor segmentation is highly desirable.

In recent years, deep learning has shown outstanding performance in liver and tumor segmentation [1]. The best performing methods are usually based on U-net [2] architecture with residual connections [3]. For the training of U-net framework, the cascade approach is commonly applied, which firstly segment the liver and then segment the tumors inside the liver. Christ et al. [4] applied two cascade U-net models for liver and tumor segmentation, respectively. The segmentation output was further refined based on 3D Conditional Random Field (3D-CRF). Chlebus et al. [5] employed two cascade models for tumor segmentation, which are followed by an object-based post-processing step. In Bellver et al. [6], the first network focuses on the liver regions, then an independent detector localizes the tumors in the liver, and finally the tumor segmentation is performed based on the localizations. They also used 3D-CRF for post-processing. Han [7], the winner of the first round of 2017 MICCAI Liver Tumor Segmentation (LiTS) challenge, developed two cascade networks working in 2.5D for the liver and tumor segmentation, respectively. Li et al. [8] proposed a novel hybrid densely connected U-net, called H-DenseUNet. A 2D DenseUNet is used to efficiently extract intra-slice features, then a 3D counterpart is employed to hierarchically aggregate volumetric contexts. Jiang et al. [9] proposed a cascade model composed of three networks: the liver localization network, the liver segmentation network, and the tumor segmentation network.

Despite their success, the above-mentioned cascade models are still struggling in small tumor segmentation. First, the size of tumors is much smaller than that of whole CT volumes as well as that of liver regions. To segment the tumor from the liver region is still like finding a needle in a haystack. Second, the cascade approaches for tumor segmentation rely on the pre-segmented liver mask. Except for the incorrect liver segmentation results will deteriorate the tumor segmentation, another defect is that the cases without liver label in the training cannot be handled. Third, adding the post-processing [4] or an additional detector [6] to the two-cascade models [7], or extending the two-cascade model to the three-cascade one [9] cannot effectively improve the accuracy of liver tumor segmentation. Training the network to better capture the essence of tumor-specific features may help to handle this small object segmentation problem.

This paper proposes a new three-stage curriculum learning

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

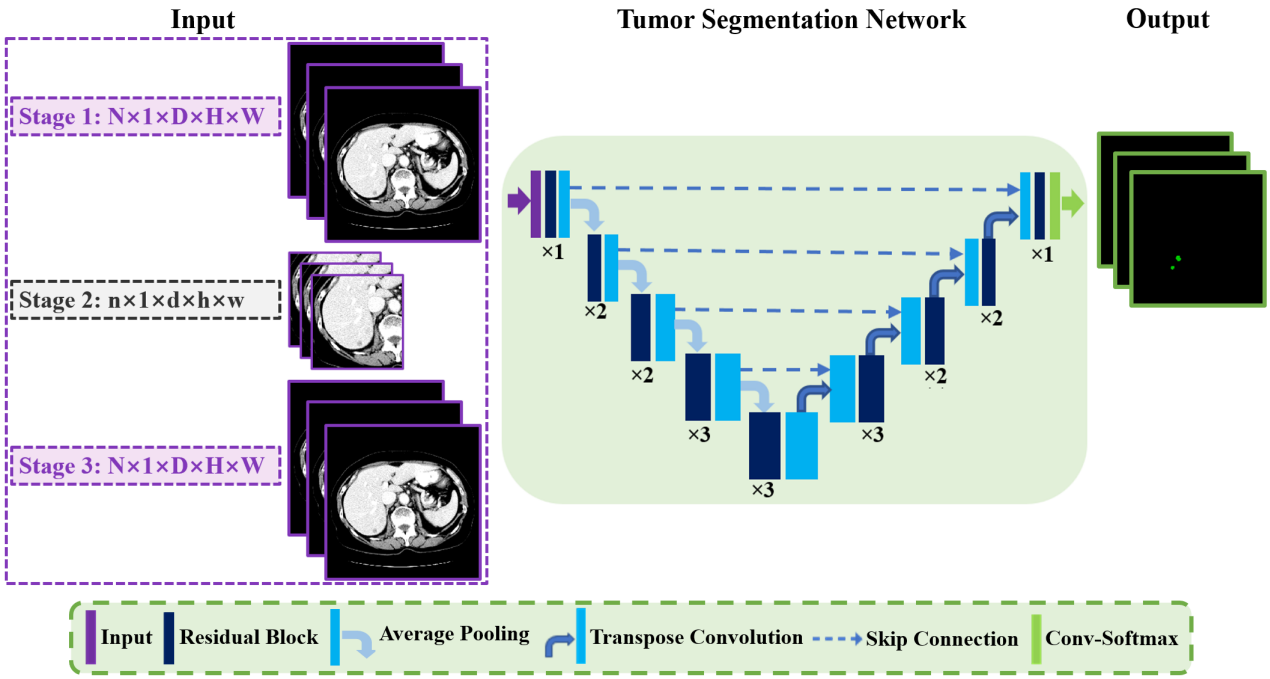

Fig. 1. An illustration of the three-stage curriculum learning approach with different scale of inputs. Network architecture: U-net based tumor segmentation network, where "× X" denotes that a residual block is repeated X times.

approach to tackle the problem of small object segmentation like liver tumor segmentation. Curriculum learning, which first proposed by [10], introduces different concepts at different times to guide the learning process. The proposed approach trains the network on two scales: tumor patches and whole input volumes. The training on tumor patches helps the model better capture specific features of tumors, while the training on the whole input volume makes the tumor-specific features effectively embedding into the global context. We integrate the two scales of training in a three-stage schema, which starts training on the whole input volume; then continues training on the tumor patches; and finally retrains on the whole input volume. In this way we can deal with the dilemma that the interested small object is drowned in the huge background. Fig. 1 illustrates our approach, which will be explained in the next section.

Our contributions are summarized as follows:

1) A new three-stage curriculum learning approach is presented to tackle the problem of small object segmentation and applied to liver tumor segmentation.
2) To the best of our knowledge, this is the first introduction of curriculum learning strategy for liver tumor segmentation.
3) As far as we know, it is also the first work based on deep learning to segment the tumor directly without the need of liver segmentation.

II. Approach

The proposed approach to tumor segmentation is illustrated in Fig. 1, which is composed of a U-net based deep network and our three-stage curriculum learning approach for training this network. The details are given as follows.

A. Tumor Segmentation Network

Recently, successful works for 3D medical image segmentation are mainly based on plain U-net architecture with residual connections [11], so we also apply this architecture in this work. However, we only employ one single network to segment the tumor directly, instead of segmenting the tumor from the liver as in commonly used cascade approaches.

The tumor segmentation network consists of an encoder and a decoder symmetrically. The encoder consists of repeated down-sampling operations which halves the size of feature map at each block. The decoder contains a set of up-sampling operations which reconstructs the feature maps in a coarse-to-fine manner. The skip connections between the encoder layers and the corresponding decoder layers enable the network to learn multi-scale semantic features. As the encoder goes deeper, the contextual information is coarser and the learned features are more related to semantics. To take full advantage of the semantic features, we carefully design the numbers of residual blocks for different layers of the encoder and decoder, which are given in Fig. 1.

B. Three-stage Curriculum Learning

Since only a very small fraction of voxels belong to a tumor in both the whole input volume and the liver region, the commonly used cascade approaches seem as finding a needle in a haystack and could miss some specific features of tumors. Moreover, the undesirable liver segmentation result hinders the performance of subsequent tumor segmentation.

Based on the thoughts above, we propose the three-stage curriculum learning approach to explore tumor-specific features more sufficiently and use it to segment the tumor directly on the CT volumes. As shown in Fig. 1, these three stages of learning are performed sequentially. The process and the role of each stage are described as follows.

• Stage 1: We start to learn the network for segmenting the tumor from the whole input 3D volumes. Instead of starting the learning process on the tumor patches, starting the learning process in this way can improve the accuracy of the final learning result, which can be seen in Section III-B.

• Stage 2: In this stage, we refine the model on 3D tumor patches extracted from the whole input volumes. The size of tumor patch is decided to be the maximum one for all the tumors in the training dataset. Except the positive samples of tumor regions, we also cropped the negative samples with the same size, which do not contain any tumor. Since tumor regions become the major part of such type of training samples, the tumor-specific features can be probed sufficiently and strengthened in the segmentation network.

• Stage 3: The training process of this stage is same as that of the first stage, but starting from the model outputted from Stage 2. This stage aims at transferring the knowledge learned from the tumor patches to the whole input volumes, in order that the tumor-specific features and the global context can be integrated sufficiently under our final purpose of segmenting the tumor from the whole input volume.

In the three-stage curriculum learning, we use the weighted Dice loss to train the network. Let p^c and g^c be the predicted probability and the ground truth probability belonging to class c (background or tumor) for an input volume, respectively; w^c be be the weighting parameter that can help alleviate the imbalance problem between the numbers of positive and negative voxels; ε be a very small number to prevent the denominator being zero, then we have

$$L_{Dice} = 1 - \sum_{c=1}^{C} w^c \times \frac{2 \times (p^c \times g^c) + \varepsilon}{p^c + g^c + \varepsilon}, \tag{1}$$

where

$$w^c = (\sum_{i \neq c}^{C} g^i + \varepsilon) / (\sum_{i=1}^{C} g^i + \varepsilon). \tag{2}$$

III. Experiments

A. Experimental Setup

Dataset and Preprocessing. Our proposed method is evaluated on two datasets. One is our clinical tumor dataset collected from Cancer Hospital and Chinese Academy of Medical Sciences. It contains 137 cases of Contrast-Enhanced CT (CECT) with arterial phase, portal venous phase and delay phase. The axial slices of all scans have the same in plane resolution of 512 × 512, but the number of slices in each scan differs among different modalities. The dataset contains the manually segmented tumors and the final annotation was validated by a senior radiologist with 15-years' experience in abdominal imaging.

Another dataset is the public MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge dataset [1]. It contains 130 CT scans for training and 70 CT scans for testing, which have the same resolution of 512×512 pixels but with different numbers of axial slices and slice thicknesses. The available ground truth is provided only for the training dataset.

We perform the experiments on the two datasets independently. On our clinical tumor dataset, all cases are randomly divided into two non-overlapping groups, 109/28 cases for training and testing, respectively. On LiTS dataset, 117/13 cases in the training set are randomly divided for training and validation, respectively, and 70 testing cases are used to test the approaches.

In medical image segmentation, data preprocessing is a prerequisite step for effective network training. In this work, we perform spacing interpolation, window transform, effective range extraction, and sub-image generation. After the preprocessing, the size of obtained input patches is determined as 64×256×256 for optimizing the trade-off between the available GPU memory used in the experiments and the contextual information in the input patches. Such input patches are used as whole input volumes in this paper. Our source code of these preprocessing steps is provided at https://github.com/Huiyu-Li/Preprocess-of-CT-data.

Implementation details. The tumor segmentation network is implemented with the PyTorch framework. All the models were trained from scratch, initialized with Kaiming uniform [12], and optimized by Adam. The initial learning rate was 0.001 and its decay rate was 0.1 for each subsequent stage of curriculum learning. The batch size of whole input volume is 1 and the corresponding patch size is 64×256×256 as described above. As for the size of tumor patches, it is 26×56×56 for our clinical tumor dataset and 64×190×190 for LiTS dataset. The corresponding batch sizes are 32 and 2, respectively. All the experiments are conducted on an NVIDIA 2080Ti GPU.

Evaluation Criteria. We evaluate the performance of the proposed approach using Dice Score (DS), which consists of Dice per Case (DC) and Dice Global (DG), Volumetric Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASSD), Maximum Surface Distance (MSD), and Root Means Square symmetric surface Distance (RMSD) [1]. A perfect segmentation yields 1 on DC and DG, while 0 on each of other metrics (VOE, RVD, ASSD, MSD and RMSD).

B. Experimental Results

a) The effectiveness of three-stage curriculum learning: To analyze the effectiveness of our proposed three-stage curriculum learning, we compared the performance of our approach with those from other related learning schemas under the same tumor segmentation network as well as the same experimental setup. These compared schemas in-cludes:

1) **Naïve Learning.** Only first stage training of our approach is considered in this schema, i.e. we naively train the tumor segmentation network on the whole input volume.

TABLE I
THE PERFORMANCE COMPARISONS BETWEEN OUR APPROACH AND ITS COUNTERPARTS ON OUR CLINICAL TUMOR DATASET.

Approach	DC	DG	VOE	RVD	ASSD	MSD	RMSD
Three-stage Curriculum Learning	**0.855**	**0.956**	**0.033**	**-0.033**	**0.075**	**1.367**	**0.228**
Whole-to-patch Curriculum Learning	0.798	0.899	0.096	-0.093	0.216	2.333	0.414
Patch-to-whole Curriculum Learning	0.652	0.775	0.206	-0.206	0.582	2.633	0.765
Naïve Learning	0.473	0.749	0.273	-0.273	0.846	3.117	1.037

TABLE II
THE PERFORMANCE COMPARISONS BETWEEN OUR APPROACH AND ITS COUNTERPARTS ON THE VALIDATION SUB-SET OF LITS TRAINING DATASET.

Approach	DC	DG	VOE	RVD	ASSD	MSD	RMSD
Three-stage Curriculum Learning	**0.822**	**0.955**	**0.235**	**0.237**	**2.458**	41.100	**5.149**
Whole-to-patch Curriculum Learning	0.799	0.947	0.265	0.329	2.533	47.112	5.904
Cascade Architecture	0.702	0.820	0.378	0.388	7.151	**36.055**	9.678
Patch-to-whole Curriculum Learning	0.633	0.852	0.457	0.473	7.513	46.936	11.249
Naïve Learning	0.671	0.809	0.421	0.860	5.115	37.407	7.965

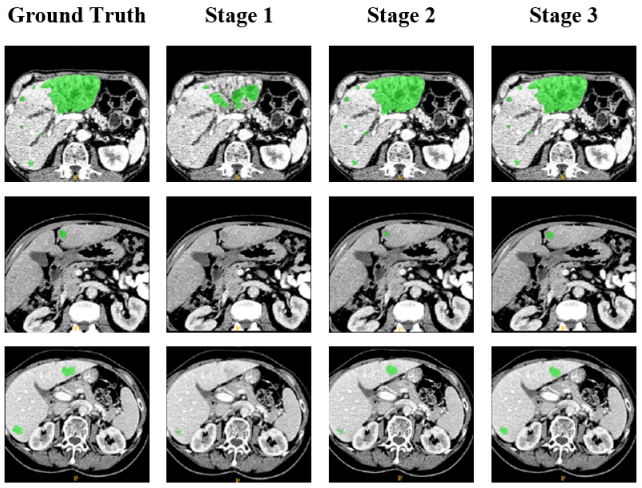

Fig. 2. Examples of liver tumor segmentation results from different training stages on the LiTS validation dataset, where each column from left to right represents ground truth, segmenta-tion results of three stages learning, respectively, and the tumors are indicted in green.

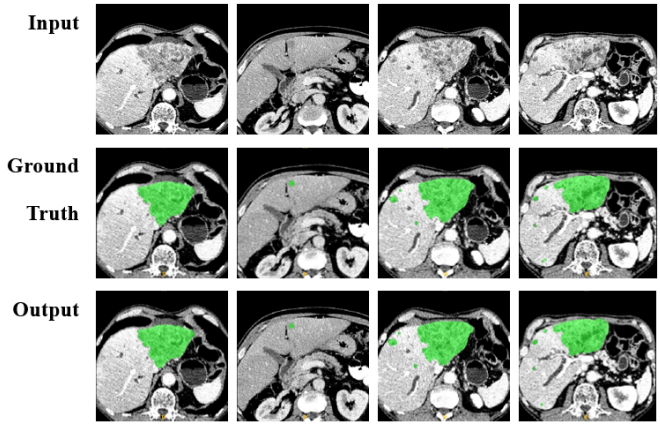

Fig. 3. Examples of final liver tumor segmentation results from the LiTS validation dataset, where each row from top to bot-tom represents input image, ground truth, and segmentation result, respectively, and the tumors are indicted in green.

2) **Whole-to-patch Curriculum Learning.** In this schema, we train the network using first two stages of our approach, starting training on the whole input volume and ending training on tumor patches.

3) **Patch-to-whole Curriculum Learning.** This schema is actually the last two stages of our approach. We start training on tumor patches and end training on the whole input volume.

4) **Cascade Architecture.** In the cascade schema, the first network is used to segment the liver and the second network is used to segment the tumor in liver. As described in section I, this is the commonly used approach in deep learning based liver tumor segmentation.

To validate the effectiveness and robustness of our approach, we conduct these experiments on our clinical tumor dataset and the public LiTS dataset. Noted that our clinical tumor dataset only contains tumor annotations, thus the cascade architecture cannot be tested on it. The comparison results of the aforementioned schemas on the clinical tumor dataset and public LiTS dataset are provided in Table I and Table II, respectively. From the results, we can conclude that:

1) Our proposed three-stage curriculum learning approach exhibits promising effectiveness for the segmentation of liver tumors. On our clinical tumor dataset, our approach achieves the best performance on all the criteria. Especially on DC and DG, two main indicators of segmentation accuracy, our approach achieves 0.855 and 0.956, respectively. On the LiTS dataset, our approach also exhibits competitive results on almost all the criteria.

2) The naive learning approach achieves poor performance, which demonstrates that learning the tumor-specific features from the whole input volume is infeasible.

3) The whole-to-patch curriculum learning exceeds the performance of the remaining three approaches. It reflects that the first two-stage of our approach is more effective than the last two stages (patch-to-whole curriculum learning). It should be noted that the last two stages of our approach have some similarities with the approach of Haarburger et al. [13], which is used to classify breast malignancy from MRI images. As the results shown in Table I and Table II, such two stages of the

TABLE III
THE PERFORMANCE COMPARISONS AMONG DIFFERENT APPROACHES ON LITS TESTING DATASET, WHERE THE APPROACHES ARE RANKED BY DC. SCORES EXCEPT OURS ARE REPORTED AS PRESENTED IN THE ORIGINAL PAPERS.

Approach	DC	DG	VOE	RVD	ASSD	MSD	RMSD
Li et al. [8]	**0.722**	0.824	0.366	4.272	1.102	6.228	1.595
Tian et al. [1]	0.702	0.794	0.394	5.921	1.189	6.682	1.726
Ours	0.690	**0.830**	0.370	**-0.052**	1.087	6.656	1.618
Li et al. [1]	0.686	0.829	0.356	5.164	1.073	**6.055**	1.562
Chlebus et al. [1]	0.676	0.796	0.383	0.464	1.143	7.322	1.728
Vorontsov et al. [1]	0.661	0.783	0.357	12.124	1.075	6.317	1.596
Yuan et al. [1]	0.657	0.820	0.378	0.288	1.151	6.269	1.678
Ma et al. [1]	0.655	0.768	0.451	5.949	1.607	9.363	2.313
Bi et al. [1]	0.645	0.735	0.356	3.431	**1.006**	6.472	**1.520**
Kaluva et al. [1]	0.640	0.770	**0.340**	0.190	1.040	7.250	1.680

patch-to-whole curriculum learning is not suitable to our liver tumor segmentation problem and leads to weak performance.

4) According to the inferior performance of patch-to-whole curriculum learning compared with three-stage curriculum learning, we can see the first learning stage provides a good start point for the last two stages.

5) Considering the inferior performance of whole-to-patch curriculum learning compared with three-stage curriculum learning, the third learning stage contributes a lot to improve the overall performance.

The examples of segmentation results in different learning stages are shown in Fig. 2. We can observe that the first learning stage can help to locate the large tumors but could miss the small tumors. The first two stages of learning can segment most of the small tumors, indicating that the tumor-specific features are effectively probed through learning on tumor patches. After the three stages of learning, we can observe that most small objects as well as large tumors can be well segmented, which further highlights the effectiveness of our proposed approach. From such visual comparison, we can claim that the performance improvement on small tumors is mainly attributed to the tumor-specific features learning. More qualitative results from our approach are visualized in Fig. 3. It shows various situations of segmentation. For the situations shown in the first two columns of the figure, there is only one tumor in each input slice but the sizes of tumors could be varied greatly across the slices. For the last two columns, there are multiple tumors in each input slice with various sizes and shapes. These cases demonstrate that the proposed approach can handle various situations and lead to satisfactory results.

B the proposeyd three-stage curriculum learning approach, we can use a single network to segment the tumor directly from the whole input volume. Compared with the commonly used cascade architecture, we can obtain obvious performance improvement, as shown in Table I and Table II. Furthermore, although such curriculum learning increases the training time, it can substantially save time for tumor segmentation during inference. In our experiments, segmenting one input volume with the size of 64×256×256 can be performed under 0.75 seconds by using the single network learned from our approach, while 2.64 seconds are needed for the two-cascade counterpart.

b) Comparison with Other Methods on LiTS testing dataset: We further compare our approach with a number of excellent competitors by submitting the result to the LiTS leaderboard. Table III shows the details. We reached a Dice per case of 0.690, Dice global of 0.830, VOE of 0.370, RVD of -0.052, ASSD of 1.087, MSD of 6.656, and RMSD of 1.618. which is a desirable performance on the LiTS challenge for tumor segmentation.

IV. CONCLUSION

In this paper, we have presented a new three-stage curriculum learning approach for handling small object segmentation and applied it to liver tumor segmentation. Compared with the commonly used cascade model, the proposed approach can lead to more precise segmentation and save computational time during inference. On our clinical tumor dataset and the LiTS dataset, the proposed approach achieves promising performance without using ensemble learning and post-processing and demonstrate the superiority over cascade counterpart. The comparison study among various combinations of learning stages further justifies the reasonability of our design. For the future work, we will validate the proposed approach on other popular medical image segmentation benchmarks, such as the kidney tumor segmentation challenge (KiTS) dataset.

ACKNOWLEDGMENT

This work was supported in part by Beijing Municipal Science and Technology Project [grant number Z181100001918002]

REFERENCES

[1] P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser *et al.*, "The liver tumor segmentation benchmark (lits)," *arXiv preprint arXiv:1901.04056*, 2019.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D Anastasi *et al.*, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 415–423.

[5] G. Chlebus, A. Schenk, J. H. Moltz, B. van Ginneken, H. K. Hahn, and H. Meine, "Automatic liver tumor segmentation in ct with fully convolutional neural networks and object-based postprocessing," *Scientific reports*, vol. 8, no. 1, pp. 1–7, 2018.

[6] M. Bellver, K.-K. Maninis, J. Pont-Tuset, X. Giró-i Nieto, J. Torres, and L. Van Gool, "Detection-aided liver lesion segmentation using deep learning," *arXiv preprint arXiv:1711.11069*, 2017.

[7] X. Han, "Automatic liver lesion segmentation using a deep convolutional neural network method," *arXiv preprint arXiv:1704.07239*, 2017.

[8] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes," *IEEE transactions on medical imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.

[9] H. Jiang, T. Shi, Z. Bai, and L. Huang, "Ahcnet: An application of attention mechanism and hybrid connection for liver tumor segmentation in ct volumes," *IEEE Access*, vol. 7, pp. 24 898–24 909, 2019.

[10] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international confer-ence on machine learning*, 2009, pp. 41–48.

[11] G. Yang, J. Gu, Y. Chen, W. Liu, L. Tang, H. Shu, and C. Toumoulin, "Automatic kidney segmentation in ct images based on multi-atlas image registration," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 5538–5541.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[13] C. Haarburger, M. Baumgartner, D. Truhn, M. Broeckmann, H. Schneider, S. Schrading, C. Kuhl, and D. Merhof, "Multi scale curriculum cnn for context-aware breast mri malignancy classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 495–503.