Deep Neural Network With Consistency Regularization of Multi-Output Channels for Improved Tumor Detection and Delineation

Hyunseok Seo, Lequan Yu, *Member, IEEE*, Hongyi Ren, Xiaomeng Li, *Member, IEEE*, Liyue Shen, and Lei Xing

***Abstract*— Deep learning is becoming an indispensable tool for imaging applications, such as image segmentation, classification, and detection. In this work, we reformulate a standard deep learning problem into a new neural network architecture with multi-output channels, which reflects different facets of the objective, and apply the deep neural network to improve the performance of image segmentation. By adding one or more interrelated auxiliary-output channels, we impose an effective consistency regularization for the main task of pixelated classification (i.e., image segmentation). Specifically, multi-output-channel consistency regularization is realized by residual learning via additive paths that connect main-output channel and auxiliary-output channels in the network. The method is evaluated on the detection and delineation of lung and liver tumors with public data. The results clearly show that multi-output-channel consistency implemented by residual learning improves the standard deep neural network. The proposed framework is quite broad and should find widespread applications in various deep learning problems.**

***Index Terms*— Artificial intelligence, cancer detection, neural networks, regularization, residual learning, segmentation.**

I. Introduction

DELINEATION via pixel-level understanding (*e.g.*, semantic segmentation) is one of the basic tasks among many applications of computer vision [1], [2] and biomedicine [3]–[9]. However, manual delineation of objects is labor-intensive, time-consuming, and suffers from inter-/intra-operator variations that often appear in radiomic features [10]–[12]. Especially, medical image analysis requires more expert-level delineation and higher coherence among the results. Therefore, significant efforts have been devoted to automate segmentation algorithms to cope with limitations of manual delineations.

Algorithms based on Deep learning (DL) have attracted much more attention in image segmentation, due to their intrinsic ability to learn complex relationships to incorporate *prior* information into network models in a data-driven manner [9], [13]. For example, Li *et al.* [14] devised a hybrid network that takes advantage of both 2D and 3D networks for liver and liver tumor segmentation in CT images. Multiple cascaded networks have been introduced for better performance [15]. Seo *et al.* [8] designed a network for liver tumor segmentation that can efficiently use object-edge information to cope with the boundary loss in the pooling operation. A modulation scheme of the loss function has been studied to handle class imbalance problems [16]. The increasing number of challenges (e.g., BraTS [17], LiTS [18], KiTS [19]) show widespread use of DL algorithms in semantic segmentation of medical images. While promising, the performance of DL-based methods is often hindered by insufficient training data or imperfect network architecture design. This situation is aggravated for medical image analysis, as the training dataset is much more limited than that for natural image applications.

In general, a neural network learns from a large set of training data under the guidance of a loss function, which drives the search for optimal network parameters by quantifying the difference between the model prediction and the ground truth. Nevertheless, minimizing a pre-defined loss function alone for a given set of training data does not always yield the optimal prediction. One of the major problems that affect the learning procedure is overfitting [20]. It has long been known as a bottleneck that degrades the performance of resultant inference model in the testing data and hinders the maximal utilization of the DL technique. Many techniques have emerged to reduce overfitting [21], such as dropout [22] and batch normalization [23]. However, further increasing the network training efficiency under limited training data remains an open problem.

Manuscript received March 31, 2021; accepted May 19, 2021. Date of publication May 28, 2021; date of current version November 30, 2021. This work was supported in part by the Korea Institute of Science and Technology under Grant 2E31122 2K02540 and Grant 2E31071, in part by the National Cancer Institute under Grant 1R01CA227713, in part by the Google Faculty Research Award the Stanford Bio-X Bowes Graduate Student Fellowship, and in part by the Human-Centered Artificial Intelligence of Stanford University. *(Corresponding author: Lei Xing.)*

Hyunseok Seo is with the Center for Bionics, Biomedical Research Institute, Korea Institute of Science and Technology (KIST), Seoul 02792, South Korea, and also with the Department of Radiation Oncology, Stanford University, Stanford, CA 94305 USA (e-mail: seo@kist.re.kr).

Lequan Yu is with the Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong (e-mail: lqyu@hku.hk).

Hongyi Ren, Liyue Shen, and Lei Xing are with the Department of Radiation Oncology, Stanford University, Stanford, CA 94305 USA (e-mail: hongyi@stanford.edu; liyues@stanford.edu; lei@stanford.edu).

Xiaomeng Li is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: eexmli@ust.hk).

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2021.3084748, provided by the authors.

Digital Object Identifier 10.1109/TMI.2021.3084748

1558-254X © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

In this paper, we design an effective network architecture to improve the network performance under limited training dataset for medical image analysis. To this end, we present an effective regularization scheme based on multi-output-channel consistency. The key idea here is to address the segmentation problem by a neural network architecture with multiple output channels, reflecting different facets of the original learning task. Specifically, we formulate the original segmentation task as the main-output channel and additionally incorporate other closely related tasks as auxiliary-output channels while maintaining the consistency between the channels. Algorithmically, it is realized by sharing the encoders and adding additive paths connecting the main-output and auxiliary-output channels in the network. By sharing the representation between the main-output channel and related auxiliary-output channels, we enable the network to learn more discriminative and generalizable features and thus achieve better performance on the original segmentation task. Notably, the proposed scheme requires no additional datasets for model training, since the labels for auxiliary outputs are derived from the input images with original labels. Recently, Zamir *et al.* [24] have presented the study of cross-task consistency at CVPR 2020, but a major difference with the proposed method is that their model concentrates on consistency constraint between the output tasks for different objectives. On the other hand, we use relevant tasks under the same objective and adopt a residual-learning-based strategy to achieve such consistency regularization. The performance of our method is demonstrated on the lung and liver tumor delineation problems. The experimental results clearly show that our network outperforms the state-of-the-art DL networks by a considerable margin, i.e., an average 10% Dice improvement in segmentation tasks.

The main contributions of this paper are summarized as follows.

- We establish an efficient DL framework for medical image segmentation tasks. Our framework improves the network performance under limited medical training data scenarios.
- We formulate the original segmentation task as by introducing a deep neural network with consistency regularization of multi-output channels (i.e., the main-output channel and additionally leverage auxiliary-output channels) for more discriminative and generalizable feature extraction. We further design additive paths connecting the main-output channel and auxiliary-output channels to adopt residual learning for multi-output channel consistency.
- Extensive experiments on two representative and challenging tumor-delineation tasks demonstrate the effectiveness of our method and outperforms the state-of-the-art methods by a large margin.

The remainders of this paper are organized as follows. We discuss the related works in Section II and elaborate the proposed framework in Section III. We present the experimental conditions and results in Section IV, and further discuss the key points of our method in Section V. Then, we draw the conclusions in Section VI.

II. Related Work

A. Medical Image Segmentation

Previous research on tumor segmentation was primarily focused on image-based modeling, which includes intensity-based thresholding [25], atlas-based models [26], deformable models [27], [28], or super-pixel method [29]. Although these approaches can produce good results, their performance depends heavily on the design or manual selections of heuristic model component(s), such as the choice of hand-crafted features (e.g, the lesion diameter and volume annotated by a radiologist) and the more robust radiomics features extracted via feature engineering [30], [31]. To incorporate statistical distribution of the patient data to improve the image segmentation problem, graph model-based methods [32], [33] were applied. Recent advances in DL tapped more potential in machine learning [34]–[37]. DL algorithms have been applied to various semantic segmentation problems, such as liver segmentation [38], [39], organ-at-risk segmentation in head and neck [40], [41], prostate segmentation [42], and brain structure and tumor segmentation [43], [44]. Most of the recent DL-based segmentation algorithms are based on U-Net architecture [45] with skip connections, e.g., dense structure [46]. It is worthwhile to note that generative adversarial network (GAN) that adversarially train two networks [47] has also been applied to image segmentation.

B. Network Regularization

DL usually provides better solution than classical algorithms for semantic segmentation. Nonetheless, it has room for improvement and there is a need to explore feature space more efficiently, especially when the network contains too many parameters to be optimized or when it is trained with insufficient training data. Several advanced network regularization techniques have been developed for improved learning, such as dropout [22] and batch or group normalization [23], [48]. Regularization on the loss function during the optimization process has also been sought after [49], [50]. Recently, shake-out [51] extended the dropout regularization and achieved a slight gain via a careful hyper-parameter selection. Regularization by latent space e.g., the least absolute shrinkage and selection operator (LASSO)-based algorithms has also been investigated [52], [53]. Finally, prior information can also be utilize to regularize the deep neural network [54]. These regularization mechanisms steadily improve the network performance on various image recognition tasks.

III. Method

A. Overview

To conduct tumor detection and delineation, in general, the network can directly predict the tumor binary masks from the input image. However, only predicting the binary mask of tumors may not always produce the optimal results, as the network could be biased to that task. The semantic segmentation typically requires visual expression to show the results of pixelated classification. Finding the binary mask is one way to proceed. In reality, other representations of the pixelated

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

classification may exist. Including these representations by introducing auxiliary tasks related to the original one would provide guidance in learning and ease the model training (i.e., regularization). Here, we propose to take advantage of different facets of the original tumor delineation task with multiple output channels via residual learning scheme. Figure 1 is an overview of our proposed consistency regularization network. Besides predicting the tumor binary masks in the main-output channel, the network also incorporates several auxiliary-output channels for relative prediction tasks. These auxiliary-output channels restore the information of input images in different ways (e.g., the original input image or clustered input images) as well as output the tumor delineation. Both the main and auxiliary-output channels are incorporated in the end-to-end learning so that each output channel is able to utilize the multiple-output-channel consistency to facilitate their predictions.

In the following subsections, we elaborate (1) how to incorporate the multi-output-channel consistency into network training, and (2) how to efficiently combine the information acquired from multiple output channels in the feature space. In our network, the encoding parts of each output channel are shared [55], [56], which extracts representative and generalizable features from multiple output channels. In addition, each output channel is bound by skip connections, i.e., additive paths. The skip connections between the different outputs of networks ease the residual learning to achieve multi-output-channel consistency.

B. Residual Learning for Multi-Output-Channel Consistency

Residual learning has been studied previously [57], [58]. The introduction of a skipping path in the residual learning simplifies the network architecture and reduces the need for training data. Here, we adopt the residual learning of correlated multiple tasks that can effectively reduce the search range in the feature space. In our proposed network the consistency regularization scheme is implemented by the additive paths among relevant multiple-output tasks. Assume that $f_z(\mathbf{x};\boldsymbol{\theta})$ is the parameterized function to map the input vector $\mathbf{x}$ to right before the additive paths in each output channel. Here, $z \in \{0, 1, \cdots c_0 - 1\}$, c_0 is the number of output channels or tasks taken in account, and $\boldsymbol{\theta}$ is the network parameters. Given the $f_z(\mathbf{x};\boldsymbol{\theta})$, we can define the output (prediction) of each channel as follows,

$$\mathbf{P}_0 = f_0(\mathbf{x};\boldsymbol{\theta}) \text{ for the main-output channel,} \tag{1}$$

$$\mathbf{P}_z = \mathbf{P}_0 + f_z(\mathbf{x};\boldsymbol{\theta})|_{z\neq 0} \text{ for the auxiliary-output channels.} \tag{2}$$

Using Eqs. (1) and (2), $\mathbf{P}_0$ for the main-output channel can be rewritten as follows,

$$\mathbf{P}_0 = \frac{1}{c_0}[\mathbf{P}_0 + \sum\nolimits_{z=1}^{c_0-1}\{\mathbf{P}_z - f_z(\mathbf{x};\boldsymbol{\theta})\}] \times \text{ for the main-output channel.} \tag{3}$$

Eq. (3) suggests that the residual learning for $\mathbf{P}_0$ can be reached by the main-output channel ($\mathbf{P}_0$) itself. Therefore, all output channels in our model conduct the residual learning for $\mathbf{P}_0$, leading to refined $\mathbf{P}_0$. In other words, residual learning is an effective way to utilize multi-output consistency. In section VII (APPENDIX), the mathematical approach and associated optimization process for the proposed network is explained in detail.

C. Network Architecture and Training Details

A state-of-the-art segmentation architecture, mU-Net [8], was applied as the backbone network. To increase the network capacitance, the feature encoder for all output channels were shared, as shown in Fig. 2. We applied three different output types for tumor segmentation with three different decoding paths in our framework.

As shown in Fig. 3, the main-output prediction p_0 is the binary mask of tumors (i.e., tumor delineation with binary masking); the auxiliary-output prediction p_1 is the combination of tumor-contour delineation and the original input image restoration; and the auxiliary-output prediction p_2 is the tumor delineation with intensity-based input-image clustering. In other words, p_2 has the simple structure information of the image and tumor-mask information.

The convolution kernels were initialized using a truncated normal distribution with mean of zero and standard deviation of 0.05, and constant bias values of 0.1. The parameters were updated by the adaptive moment (Adam) algorithm [59] with an adaptive learning rate to improve learning efficiency. The starting learning rate was empirically set as 0.001 to avoid divergence and improve convergence speed, and it was scaled by 0.97 for every 5 epochs. The decay of moving average for batch normalization was set to 0.9. The probability of dropout for regularization was set to 0.65. The batch size was set as 15 to balance the GPU memory constraints and learning time. The samples were shuffled in each training epoch.

We referred Myronenko's study [55] to apply multiple loss functions and set weights for them. This study provided the analysis of multiple loss functions and their scaling factors, which is relevant to multi-task learning (MTL). We set loss functions for each output channel and the corresponding scaling factors as follows: dice loss for p_0, combination of L_2 loss and KL loss for p_1. The L_1 loss were applied to p_2, as the prediction result for the second auxiliary-output channel includes two clustered regions and L_1 loss works well for simple texture region [60] via imposing sparsity on loss calculation. The weights for the three output channels were initialized as $\omega_0 : \omega_1 : \omega_2 = 1 : 0.1 : 0.1$. Finally, the network was trained to minimize the total value of these loss functions. Furthermore, additional forward calculation was performed to check the feasibility of adaptive update for the weights in each loss function. Specifically, weights were adaptively updated by the rule in Fig. 4 according to the specific performance of three tasks at the current iteration. The weights were adjusted by the deterministic ratio at every iteration to get the best dice score for the prediction result of the main-output channel at every iteration. The network optimization was performed on a DGX Station from NVIDIA running Linux operating system with an Intel Xeon E5-2698 v4 2.2 GHz (20-Core) CPU and two Tesla V100 GPUs (32 GB memory for

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

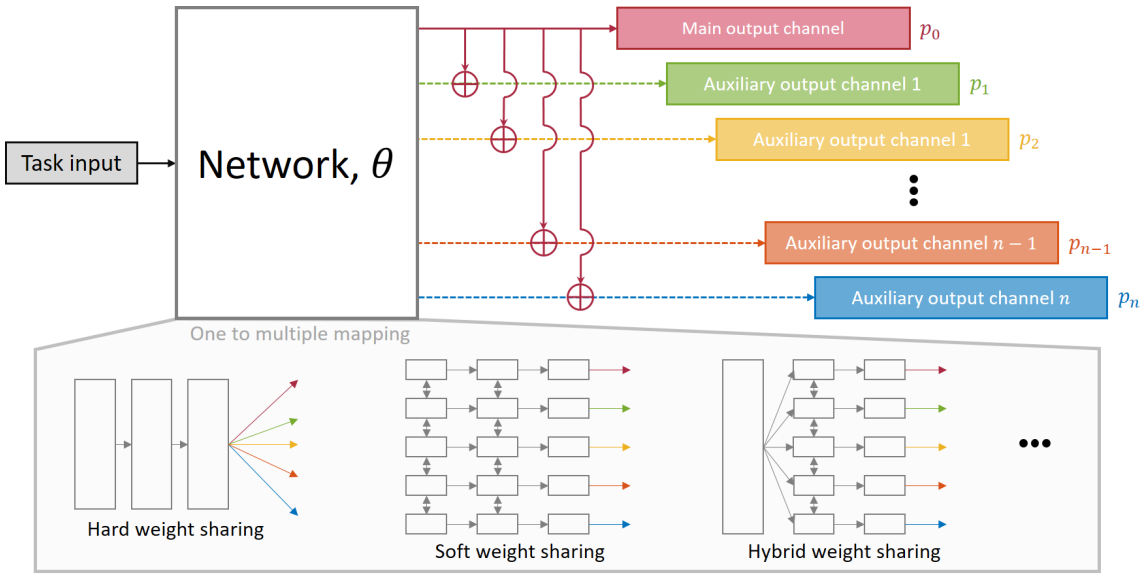

Fig. 1. Illustration of multi-output-channel consistency scheme. The whole framework is composed of different but related output channels to guide network learning.

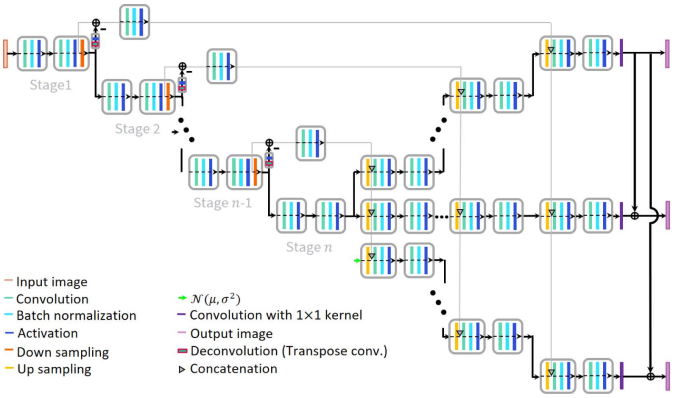

Fig. 2. Multi-output-channel consistency regularized deep neural network.

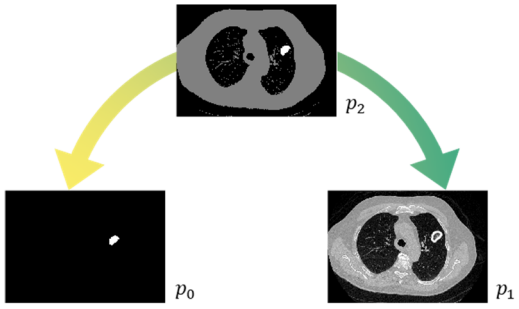

Fig. 3. Various segmentation tasks applied to the proposed network.

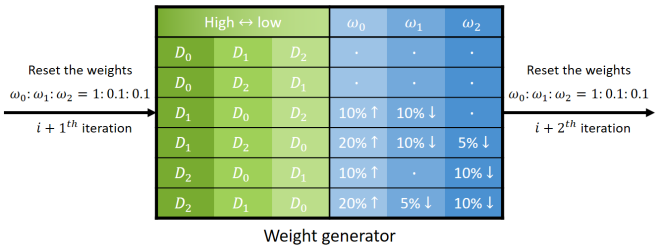

Fig. 4. Illustration of weight generator for adaptive weights. First, the weights are reset before generating the new weights at each iteration. Then, dice scores of D_0, D_1, and D_2 at every iteration are calculated from the main-output channel, auxiliary-output channel 1, and auxiliary-output channel 2, respectively. According to the relative comparison among dice scores, weights are changed to get the highest dice score for the main-output channel. *e.g.*, if D_0 is highest, all weights are fixed, and if D_0 is not highest, w_0 should be increased. The rates of change were determined empirically.

TABLE I
THE QUANTITATIVE RESULTS OF THE PROPOSED NETWORKS AND OTHER COMPARED NETWORK FOR THE LUNG-TUMOR DATASETS (AVERAGED PRECISION, RECALL, AND DICE SCORE)

Network	Precision	Recall	Dice score
mU-Net	0.7941	0.6729	0.6776
Ensemble	0.8006	0.6706	0.6754
MTL(F)	0.8259	0.7192	0.7222
MTL(A)	0.8183	0.7319	0.7400
Proposed(F)	0.8727	0.8062	0.8279
Proposed(A)	0.8873	0.8227	0.8429

each GPU). The network architecture was implemented with the well-known DL framework TensorFlow [61]. The expanded network architecture to multi-class segmentation dataset is shown in Supplementary Figure 1(a)[1] and to fully 3D model is shown in Supplementary Figure 1(b).[2]

[1]Supplementary materials are available in the supporting documents.
[2]Supplementary materials are available in the supporting documents.

IV. EXPERIMENTS

A. Datasets and Preparation

The public dataset for the lung tumor segmentation in this study were obtained from the Decathlon Challenge [62]. The dataset includes 60 CT scans with small tumors. The image size is 512 × 512. In our study, we first randomly selected 48 patient scans for training, 4 for validation, and

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

TABLE II
THE 95 % CONFIDENCE INTERVAL FOR THE RESULTS IN TABLE I

Network	Precision	Recall	Dice score
mU-Net	0.00128	0.01398	0.01327
Ensemble	0.00136	0.01306	0.01220
MTL(F)	0.00121	0.01332	0.01240
MTL(A)	0.00098	0.01220	0.00878
Proposed(F)	0.00086	0.00743	0.00556
Proposed(A)	0.00081	0.00722	0.00528

TABLE III
THE *p*-VALUES FOR THE PROPOSED METHOD WITH ADAPTIVE WEIGHTS (I.E., PROPOSED (A)) OF THE LUNG DATASET. WE PERFORMED *t*-TEST UNDER THE NULL HYPOTHESIS H0: $\mu_{\text{PROPOSED (A)}} = \mu_C$, WHERE C STANDS FOR THE COMPARED METHODS AND μ IS THE MEAN VALUES IN TABLE I. WE CAN REJECT H0 AT THE SIGNIFICANCE LEVEL 0.05 BECAUSE ALL *p*-VALUES ARE FOUND TO BE LESS THAN 0.05. THE PROPOSED (A) DO NOT HAVE *p*-VALUES BECAUSE THE *t*-TESTS WERE PERFORMED BASED ON HOW MUCH THE RESULTS OF OTHER METHODS WERE DIFFERENT FROM THAT OF THE PROPOSED (A)

Network	Precision	Recall	Dice score
mU-Net	<0.00001	<0.00001	<0.00001
Ensemble	<0.00001	<0.00001	<0.00001
MTL(F)	<0.00001	<0.00001	<0.00001
MTL(A)	<0.00001	<0.00001	<0.00001
Proposed(F)	<0.00001	0.00143	0.00009
Proposed(A)	-	-	-

8 for test. We then repeated the process with a different patient-level split of training, validation, and test datasets. The final results were obtained by averaging test results from five repetitions of data splits. The data-split policy is described in Supplementary Figure 2.[3] The method was also applied to the liver tumor segmentation with datasets from the Liver Tumor Segmentation Challenge (LiTS-ISBI2017) [18]. The dataset includes 130 abdomen contrast CT scans. The image size of each CT slice is also 512 × 512. Total 104 patient scans were used for training, 5 for validation, and 21 for test. Again, a patient-level split was performed in the same way as described earlier. To train the network, the original tumor-annotation mask images were regenerated to the annotations for the two different auxiliary-output channels, as shown in Fig. 3. The generated label images for auxiliary-output channels were based on the intensity level of the input and the corresponded label pair of the datasets. Supplementary Figure 3[4] describes the ways of each annotation is generated.

B. Performance Evaluation

To verify the performance of the proposed method, we compare the proposed method with three other different methods. The first case was the original mU-Net for only predicting p_0. The second case was the (independent/separate training models) ensemble learning. Specifically, we trained one mU-Net for only predicting p_0, another mU-Net for only predicting p_1, and the other mU-Net for only predicting p_2. At the final stage, the segmentation region was obtained by averaging the tumor masks from the three channel predictions. The third compared model (MTL) was the network with the encoder sharing from multiple-output channels, but without the additive paths, so that it is similar with the previous Myronenko's method [55] and can be used to validate the effect of the additive paths in the network. Also, we conduct ablative study on our methods with and without the proposed adaptive weight adjusting scheme in Section IV-C,D (denoted as P(A), and P(F), respectively). For the mU-Net in the first case, the loss function was defined as Dice loss. For the ensemble learning, the mU-Nets employed Dice loss for p_0, (L_2 loss + KL loss) for p_1, and L_1 loss for p_2, respectively. For the MTL case, the same loss functions with the proposed method were applied. The evaluations were performed with the widely used metrics of precision, recall, and dice score for the prediction results. Here, MTL and proposed method used only p_0 for evaluation. All methods adopted the same dropout and batch normalization scheme. In addition, the performance was also analyzed with respect to the size of training datasets and tumor sizes to show the effectiveness of the proposed network. All processing for data analysis were implemented using MATLAB (9.7.0.1261785, R2019b, The MathWorks Inc., Natick, MA).

C. Results on Lung Datasets

The averaged precision, recall, and dice score of lung datasets are shown in Tables I-III and Supplementary Figure 4.[5] Precisions of all methods are more than 0.96, which means the detected tumors are well delineated. Furthermore, the proposed framework has the highest precision. As for the recall metric, the compared networks have values lower than 0.86, while the proposed network achieves 0.91 and 0.92 for fixed weights (P(F)) and adaptive weights (P(A)), respectively, suggesting that the proposed network has less chance to fail the tumor delineation. Note that the recall can vary up to 9 % depending on the different learning schemes. The proposed network produces the highest dice score (0.93) and the adaptive weighting manner has a slightly higher improvement. We show the distribution of the dice score across different tumor sizes in Fig. 5. This distribution shows the reason why the proposed network achieves a higher recall scores than other networks. As we can see, for the small size of tumor targets less than 30 pixels, all methods fail the target delineation. However, for the relatively larger tumor targets, only the proposed network successfully segments the tumors while the compared networks fail to locate the tumors less than 70 pixels. The dice scores are almost saturated when the size of tumors is larger than 100 pixels. The other dice score plot with respect to the size of datasets in right side of Fig. 5 shows the performance of the proposed network under insufficient training datasets. When only 70 % training datasets were applied, the proposed network still has a dice score higher than 0.8, while dice scores of other compared methods are around 0.7. The dice score of MTL(F) (M) is always higher than those of mU-Net (U) and Ensemble (E) cases when the available datasets are reduced.

[3]Supplementary materials are available in the supporting documents.
[4]Supplementary materials are available in the supporting documents.
[5]Supplementary materials are available in the supporting documents.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

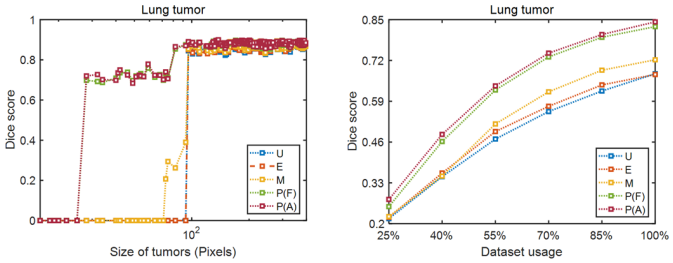

Fig. 5. Dice scores with respect to the size of lung tumors and datasets are shown in the second row. U, E, M, P(F), and P(A) mean the mU-net, ensemble learning, multi-task leaning, proposed network with fixed weights, and the proposed network with adaptive weights, respectively.

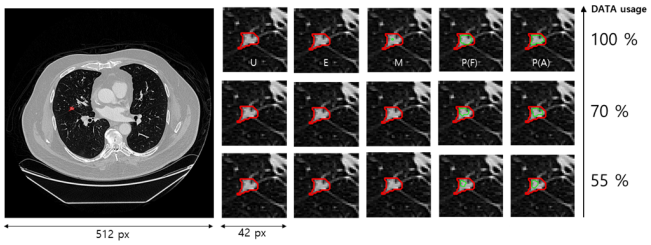

Fig. 6. Segmentation images with respect to the size of training datasets for lung-tumor datasets. The first column in the magnified images (42 px) is the results of mU-Net (U). The second, third, fourth, and fifth columns are the results of Ensemble (E), MTL (M), proposed network with fixed weights P(F) and proposed network with adaptive weights P(A), respectively. The red contours denote the ground truth and the green contours represent the prediction results from each method. No green contours in the magnified images means the method fail to delineate the tumor.

TABLE IV
THE QUANTITATIVE RESULTS OF THE PROPOSED NETWORKS AND OTHER COMPARED NETWORK FOR THE LIVER-TUMOR DATASETS (AVERAGED PRECISION, RECALL, AND DICE SCORE)

Network	Precision	Recall	Dice score
mU-Net	0.8309	0.7087	0.7419
Ensemble	0.8307	0.7126	0.7422
MTL(F)	0.8557	0.7418	0.7755
MTL(A)	0.8593	0.7575	0.7927
Proposed(F)	0.8992	0.8070	0.8424
Proposed(A)	0.9057	0.8178	0.8513

Figure 6 shows some visual segmentation results with respect to different sizes of the training dataset. As can be observed from Fig. 6, when training with small size of the datasets, all networks fail delineation of the small target tumor except the proposed network. The proposed network can successfully detect and segment the tumors even if only half of training datasets were applied.

We can also see the effect of gain from the main-output channel and the auxiliary-output channels respectively, as shown in Fig. 7 (left side). The binary mask (p_0) served as the main-output channel provides the highest dice score. Figure 7 (right side) also shows the relative computational costs of each method. Although the proposed network has twice larger computing cost than that of mU-Net, it has the highest accuracy with smaller cost than Ensemble learning and similar cost with MTL.

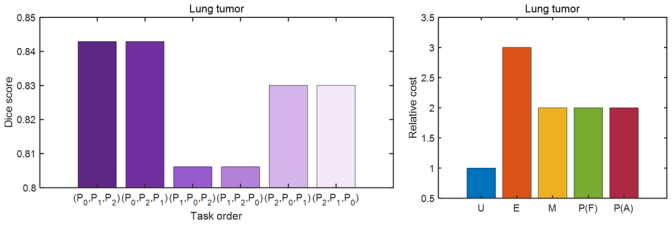

Fig. 7. Performance of the proposed method with respect to order of the output channels in P(A) case (left side). For example, (p_0, p_1, p_2) means p_0 for main-output channel, p_1 for sub-output channel 1, and p_2 for sub-output channel 2. In the same manner, (p_0, p_2, p_1) means p_0 for main-output channel, p_2 for sub-output channel 1, and p_1 for sub-output channel 2. Relative cost for computing for each method (Right). The proposed network has higher cost than mU-Net case but not higher than those of other methods.

TABLE V
THE 95 % CONFIDENCE INTERVAL FOR THE RESULTS IN TABLE IV

Network	Precision	Recall	Dice score
mU-Net	0.00020	0.02576	0.02506
Ensemble	0.00014	0.02493	0.02392
MTL(F)	0.00014	0.02338	0.02230
MTL(A)	0.00014	0.01316	0.01153
Proposed(F)	0.00015	0.00625	0.00520
Proposed(A)	0.00012	0.00600	0.00481

TABLE VI
THE p-VALUES FOR THE PROPOSED METHOD WITH ADAPTIVE WEIGHTS (I.E., PROPOSED (A)) OF THE LIVER DATASET. WE PERFORMED t-TEST UNDER THE NULL HYPOTHESIS H0: $\mu_{\text{PROPOSED (A)}} = \mu_C$, WHERE C STANDS FOR THE COMPARED METHODS AND μ IS THE MEAN VALUES IN TABLE IV. WE CAN REJECT H0 AT THE SIGNIFICANCE LEVEL 0.05 BECAUSE ALL p-VALUES ARE FOUND TO BE LESS THAN 0.05. THE PROPOSED (A) DO NOT HAVE p-VALUES BECAUSE THE t-TESTS WERE PERFORMED BASED ON HOW MUCH THE RESULTS OF OTHER METHODS WERE DIFFERENT FROM THAT OF THE PROPOSED (A)

Network	Precision	Recall	Dice score
mU-Net	<0.00001	<0.00001	<0.00001
Ensemble	<0.00001	<0.00001	<0.00001
MTL(F)	<0.00001	<0.00001	<0.00001
MTL(A)	<0.00001	<0.00001	<0.00001
Proposed(F)	<0.00001	0.01271	0.01232
Proposed(A)	-	-	-

D. Results on Liver Datasets

The averaged precision, recall, and dice score of different methods on liver datasets are presented in Tables IV-VI and Supplementary Figure 4.[6] In this case, precisions were not sensitive to each network. All networks produce more than 0.99 values on the precision, which implies that the delineated tumors are extremely accurate. For the recall scores, the similar trend to lung tumor segmentation is observed. The proposed network has the highest recall value of 0.89 and 0.90 for P(F) and adaptive weights P(A), respectively, while other compared networks have values lower than 0.87. In other words, the proposed network hardly fails the tumor delineation in comparison to other networks. Besides precisions

[6]Supplementary materials are available in the supporting documents.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

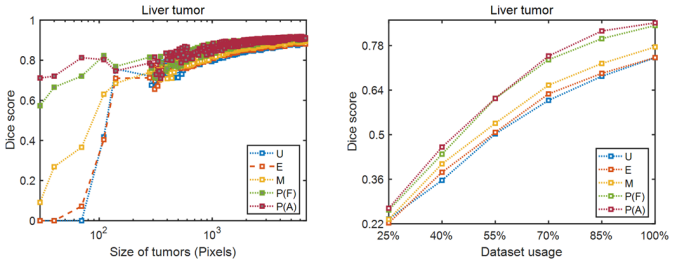

Fig. 8. Dice scores with respect to the size of liver tumors. Here, 100 pixels and 1000 pixels correspond to 32 mm^2 and 336 mm^2, respectively. U, E, M, P(F), and P(A) mean the mU-net, ensemble learning, multi-task leaning, proposed network with fixed weights, and the proposed network with adaptive weights, respectively.

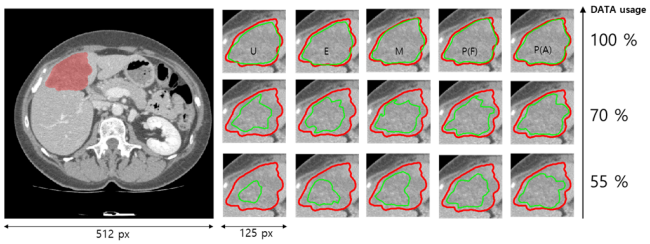

Fig. 9. Segmentation images with respect to the size of training datasets for liver-tumor datasets. The full length 512 pixels of the CT scan corresponds to 30cm, the magnification window length 125px corresponds to 73mm. The first column in the magnified images (125 px) is the results of mU-Net (U). The second, third, fourth, and fifth columns are the results of Ensemble (E), MTL (M), proposed network with fixed weights P(F) and proposed network with adaptive weights P(A), respectively. The red contours denote the ground truth and the green contours represent the prediction results from each method. No green contours in the magnified images means the method fail to delineate the tumor.

and recalls, the proposed network also achieves the highest dice score. On the other hands, the compared networks have dice score lower than 0.92. There are few small sizes of tumors in liver datasets, but the proposed network still achieves higher dice score across almost different size of tumors (see Fig. 8). For liver datasets, the plot of dice score with respect to different size of training datasets also shows that the proposed network outperforms other networks (Fig. 8). The visual segmentation results of different methods are shown in Fig. 9. The delineations for each network are pretty accurate, as the liver tumors are more obvious and larger than that of lung tumors, when full training datasets were applied. However, the accuracy gets worse when training datasets are decreased, while the proposed network provides more accurate delineations than other networks under the limited training dataset.

V. Discussion

Although DL has achieved dramatically better performance than conventional machine learning models for many medical image analysis problems, there is still room for improvement. A bottleneck issue impeding the widespread applications of DL models is that the training process of deep neural networks is vulnerable to insufficient training data and the small tumor targets. This challenging issue can be alleviated by imposing regularization methods in the optimized objective. Batch or group normalization are widely used regularization techniques. Shibani *et al.* [63] have shown that real impact of the batch normalization is not internal covariance shift but smoothing of landscape in the underlying optimization problem. However, it may not work well in some scenarios (e.g., small size of mini-batch and large variance of the dataset) [64]. Other regularization methods may be directly applied to the objective functions, but success has been limited because they usually rely on some prerequisites for regularization terms that are not easy to expand to general applications. In this study, we bring up a new network regularization scheme based on multi-output-channel-consistency learning.

In the proposed regularization scheme, the output channels are discomposed into two types: one is a main-output channel and the other are auxiliary-output channels relevant to the main task. Moreover, there are additive paths connecting the main-output channel and the auxiliary-output channels for efficient joint learning. Through the additive paths, the regularization is achieved by interaction between the main-output channel and the auxiliary-output channels. The main-output channel and auxiliary-output channels regularize mutually during the network update process. In other words, in our learning model, multiple auxiliary-output channels can provide different facets of the inferred information, so that the network learning can effectively utilize multi-output-task consistency via residual learning. The residual images of the multiple outputs in our method are likely to improve the learning efficiency via the task consistency, as shown in Supplementary Figure 5.[7] The grad-CAM analysis based on [66] also suggests that the proposed network generates more discriminative representations to better describe the target tumor as compared to other methods (Supplementary Figure 6[8]).

Further performance improvement may be possible by adding more auxiliary-output channels until the information from auxiliary-output channels becomes redundant. However, due to the GPU memory limitation, we used the main-output channel and two auxiliary-output channels in our experiments. Notably, the proposed network is capable of delineating small tumors less than 100 pixels even with only 55 % of the training datasets. Furthermore, using the adaptive weighting, we can get slightly higher accurate delineations. The weight for each of the training steps is shown in Supplementary Figure 7[9] and more sophisticated weight policies can provide further improvement of the network performance. If we design the adaptive update rule more carefully, a better result would be expected. Training with a small sample size often causes overfitting to the dataset. To minimize the limitations caused by small data samples and ensure generality of this study, we repeated the learning process five times with different splits of training, validation, and test datasets (i.e., patient-level splits). We also expanded our network to multi-class segmentation dataset (BraTS). The performance of the proposed method is clearly better as compared with other methods

[7]Supplementary materials are available in the supporting documents.

[8]Supplementary materials are available in the supporting documents.

[9]Supplementary materials are available in the supporting documents.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

without consistency regularization of multi-output channels (Supplementary Table I[10]). In the future work, we would also investigate how to create a more efficient auxiliary output channel (task) and how to reduce the empirical choice for the adaptive update.

While colossal advances have been made in using DL for medical image analysis, there is little guarantee that perfect inference will be resulted when a model is generalized to new data unseen in the training. The proposed methodology improves the robustness of DL model by leveraging the interaction of output channels applied to the network. From the results shown in this study, accuracy of the segmentation results reaches the highest values beyond limitation of other previous methods and provides new chances for other practical applications.

VI. Conclusion

This paper presents a multi-output-channel consistency regularization method for DL-based image segmentation. In the proposed strategy, the main-output channel and auxiliary-output channels are connected through the additive paths and make a joint decision with consideration of the requirement of each individual channel. The evaluation performed with public lung- and liver-tumor segmentation datasets demonstrates the superiority of the proposed method. Finally, while the current study is focused on segmentation, the proposed residual learning methodology is quite general and can be applied to other practical applications which can be formulated as a problem with multiple output channels.

Appendix

A. The Learning Process of Single Output Channel

The deep neural networks are optimized to minimize the predefined loss objective function. To find the minimum loss value with respect to the network parameter $\boldsymbol{\theta} = \left[\theta_0 \cdots \theta_{d-1}\right]^T$, the gradient descent method is usually leveraged in optimization, where the gradient of the loss function ($\nabla\mathcal{L}$) is calculated as follows,

$$\begin{aligned}\boldsymbol{\theta}^{i+1} &\rightarrow \boldsymbol{\theta}^i - \lambda\, \nabla\mathcal{L}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^i},\\ \nabla\mathcal{L} &= \nabla\mathcal{L}_0\left(\mathbf{p}_0\right)\cdot\mathbf{p}_0'\left(\boldsymbol{\theta}\right),\\ &= \left[\boldsymbol{\ell}_0^{p_0}\cdot\boldsymbol{\mathcal{P}}_0^{\theta_0}\ \cdots\ \boldsymbol{\ell}_0^{p_0}\cdot\boldsymbol{\mathcal{P}}_0^{\theta_{d-1}}\right]^{\mathrm{T}},\\ &\equiv \boldsymbol{\mathcal{V}}\left[\boldsymbol{\ell}_0^{p_0}\cdot\boldsymbol{\mathcal{P}}_0^{\theta_q}\right]_{q=0}^{d-1},\end{aligned} \tag{A.1}$$

where i is the iteration number, $\cdot$ denotes the inner product, $\mathcal{L}_0$ is the specific loss function which is a function of $\mathbf{p}_0$, $\mathbf{p}_0 = \left[x_0^{p_0}\cdots x_{r-1}^{p_0}\right]^{\mathrm{T}}$ is a vector of the prediction result at the specific network with parameters of $\boldsymbol{\theta}$, and λ is a learning rate. $\boldsymbol{\mathcal{V}}$ denotes a vector that is composed of its elements. For example, $\boldsymbol{\mathcal{V}}\left[a_j\right]_{j=0}^{b-1} = \left[a_0\cdots a_{b-1}\right]^{\mathrm{T}}$. $\boldsymbol{\ell}_0^{p_0}$ and $\boldsymbol{\mathcal{P}}_0^{\theta_q}$ are defined as $\left[\frac{\partial\mathcal{L}_0}{\partial x_0^{p_0}}\cdots\frac{\partial\mathcal{L}_0}{\partial x_{r-1}^{p_0}}\right]^{\mathrm{T}}$ and $\left[\frac{\partial x_0^{p_0}}{\partial\theta_q}\ \cdots\ \frac{\partial x_{r-1}^{p_0}}{\partial\theta_q}\right]^{\mathrm{T}}$, respectively. In other words, each step to update network weights is affected by the selected loss function L_0, and the current network prediction result $\mathbf{p}_0$.

[10]Supplementary materials are available in the supporting documents.

B. Expanded Learning From Multiple Output Channels

Eq. (A1) describes the gradient calculation of a single loss function and the general form of multiple loss elements with respect to multiple output channels can be expanded as follows,

$$\begin{aligned}\nabla\mathcal{L} &= \boldsymbol{\mathcal{V}}\left[\left(\sum\nolimits_{k=0}^{c_0-1}\alpha_{0,k}\boldsymbol{\ell}_{0,k}^{p_0}\right)\cdot\boldsymbol{\mathcal{P}}_0^{\theta_q}\right]_{q=0}^{d-1},\\ &\equiv \boldsymbol{\mathcal{V}}\left[\hat{\boldsymbol{\ell}}_0^{p_0}\cdot\boldsymbol{\mathcal{P}}_0^{\theta_q}\right]_{q=0}^{d-1},\end{aligned} \tag{A.2}$$

where c_0 is the number of the tasks taken in account. $\alpha_{0,k}$ is a weight for scaling the k-th loss elements and $\boldsymbol{\ell}_{0,k}^{p_0}$ is defined as $\left[\frac{\partial\mathcal{L}_{0,k}}{\partial x_0^{p_0}}\cdots\frac{\partial\mathcal{L}_{0,k}}{\partial x_{r-1}^{p_0}}\right]^{\mathrm{T}}$ when $\mathcal{L}_0 = \sum_{k=0}^{c_0-1}\alpha_{0,k}\mathcal{L}_{0,k}$. Then, the weight vector for $\boldsymbol{\mathcal{P}}_0^{\theta_q}$ is averaged by multiple $\boldsymbol{\ell}_{0,k}^{p_0}$ so that the updating step can be toward more precise way than the case using a single loss element (i.e., noise vector smoothing). Now, we can explain how multiple loss aggregation in single output channel contributes to network learning. In other words, intuitively, there is no one-size-fits-all loss function that can include the multidimensional information with a scalar value.

Next, the easiest way to combine multiple output channels with Eq. (A2) is to define the final loss function as weighted sum of the loss functions defined in each output channel as follows,

$$\nabla\mathcal{L} = \boldsymbol{\mathcal{V}}\left[\sum\nolimits_{k=0}^{m}\omega_k\hat{\boldsymbol{\ell}}_k^{p_k}\cdot\boldsymbol{\mathcal{P}}_k^{\theta_q}\right]_{q=0}^{d-1}. \tag{A.3}$$

Here, for k-th output channel, $\hat{\boldsymbol{\ell}}_k^{p_k}$ is defined as $\sum_{j=0}^{c_k-1}\alpha_{k,j}\boldsymbol{\ell}_{k,j}^{p_k}$, and $\boldsymbol{\mathcal{P}}_k^{\theta_q}$ is defined as $\left[\frac{\partial x_0^{p_k}}{\partial\theta_q}\ \cdots\ \frac{\partial x_{r-1}^{p_k}}{\partial\theta_q}\right]^{\mathrm{T}}$. In this case, it is expected to optimize the weighted sum of loss functions in different output channels. However, this straightforward scheme does not guarantee that predictions of all output channels can reach the optimal results at the same time.

C. Expanded Learning From Multiple Output Channels With Additive Paths

To achieve a better prediction result through the interaction among multiple output channels, in our proposed framework, the multiple output channels are categorized into two types: main-output channel (prediction result $\mathbf{p}_0$) and auxiliary-output channels (prediction results $\mathbf{p}_1, \cdots, \mathbf{p}_n$). In this way, we can focus on improving the main-output channel with the regularization from the auxiliary-output channels learning. To this end, the network outputs of different output channels are connected by additive paths from the main-output channel to other auxiliary-output channels, as shown in Fig. 1. With the additive paths, the gradient calculation of the final loss functions of the proposed network can be represented as follows,

$$\begin{aligned}\nabla\mathcal{L} = \boldsymbol{\mathcal{V}}&\left[\left(\sum\nolimits_{k=0}^{n}\omega_k\hat{\boldsymbol{\ell}}_k^{p_k}\right)\cdot\boldsymbol{\mathcal{P}}_0^{\theta_q}\right]_{q=0}^{d-1}\\ &+\boldsymbol{\mathcal{V}}\left[\sum\nolimits_{k=1}^{n}\omega_k\hat{\boldsymbol{\ell}}_k^{p_k}\cdot\left(\boldsymbol{\mathcal{P}}_k^{\theta_q}-\boldsymbol{\mathcal{P}}_0^{\theta_q}\right)\right]_{q=0}^{d-1},\end{aligned} \tag{A.4}$$

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

where $\boldsymbol{\mathcal{P}}_k^{\theta_q} - \boldsymbol{\mathcal{P}}_0^{\theta_q} = \left[\frac{\partial(x_0^{p_k} - x_0^{p_0})}{\partial \theta_q} \cdots \frac{\partial(x_{r-1}^{p_k} - x_{r-1}^{p_0})}{\partial \theta_q} \right]^{\mathrm{T}}$ means the derivative of difference in prediction results. In the proposed network, the weight for $\boldsymbol{\mathcal{P}}_0^{\theta_q}$ is averaged by $\hat{\boldsymbol{\ell}}_k^{p_k}$ (i.e., regularization of main-output channel via loss functions of multi-output channels). For the auxiliary-output channels, we employ the residual learning technique [65] as shown $(\boldsymbol{\mathcal{P}}_k^{\theta_q} - \boldsymbol{\mathcal{P}}_0^{\theta_q})$ in Eq. (A4) so that the network can keep attention on the main-output channel prediction for more efficient learning. Consequently, the auxiliary-output channels and the corresponding loss functions are able to provide a regularization effect to the main-output channel, which is relevant to the first term in Eq. (A4). Also, the main-output channel serves as an 'anchor' in the residual learning to make the learning of auxiliary-output channels easier, which is relevant to the second term in Eq. (A4).

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[3] L. Xing, M. Giger, and J. Min, *Artificial Intelligence in Medicine*. Amsterdam, The Netherlands: Elsevier, 2020.

[4] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[5] D. S. W. Ting *et al.*, "Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes," *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[6] R. Poplin *et al.*, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," *Nature Biomed. Eng.*, vol. 2, no. 3, p. 158, 2018.

[7] B. Ibragimov, D. Toesca, D. Chang, Y. Yuan, A. Koong, and L. Xing, "Development of deep neural network for individualized hepatobiliary toxicity prediction after liver SBRT," *Med. Phys.*, vol. 45, no. 10, pp. 4763–4774, Oct. 2018.

[8] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1316–1325, May 2020, doi: 10.1109/TMI.2019.2948320.

[9] H. Seo *et al.*, "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," *Med. Phys.*, vol. 47, no. 5, May 2020, doi: 10.1002/mp.13649.

[10] Q. Dou *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Med. Image Anal.*, vol. 41, pp. 40–54, Oct. 2017.

[11] H. Moradmand, S. M. R. Aghamiri, and R. Ghaderi, "Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma," *J. Appl. Clin. Med. Phys.*, vol. 21, no. 1, pp. 179–190, Jan. 2020.

[12] S. Pati *et al.*, "Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy glioblastoma atlas project (Ivy GAP) dataset," *Med. Phys.*, vol. 47, no. 12, pp. 6039–6052, Dec. 2020.

[13] L. Shen, W. Zhao, and L. Xing, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nature Biomed. Eng.*, vol. 3, no. 11, pp. 880–888, Nov. 2019.

[14] 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 Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[15] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "NnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.

[16] H. Seo, M. Bassenne, and L. Xing, "Closing the gap between deep neural network modeling and biomedical decision-making metrics in segmentation via adaptive loss functions," *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 585–593, Feb. 2021.

[17] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[18] P. Bilic *et al.*, "The liver tumor segmentation benchmark (LiTS)," 2019, *arXiv:1901.04056*. [Online]. Available: http://arxiv.org/abs/1901.04056

[19] N. Heller *et al.*, "The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes," 2019, *arXiv:1904.00445*. [Online]. Available: http://arxiv.org/abs/1904.00445

[20] J. Brownlee, *Better Deep Learning: Train Faster, Reduce Overfitting, and Make Better Predictions*. San Francisco, CA, USA: Machine Learning Mastery, 2018.

[21] X. Ying, "An overview of overfitting and its solutions," *J. Phys., Conf. Ser.*, vol. 1168, no. 2, Feb. 2019, Art. no. 022022.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[24] A. R. Zamir *et al.*, "Robust learning through cross-task consistency," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11197–11206.

[25] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. Technol.*, vol. 19, no. 1, pp. e252–e258, Feb. 2013.

[26] D. Li *et al.*, "Augmenting atlas-based liver segmentation for radiotherapy treatment planning by incorporating image features proximal to the atlas contours," *Phys. Med. Biol.*, vol. 62, no. 1, p. 272, 2016.

[27] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang, "Automatic liver segmentation based on shape constraints and deformable graph cut in CT images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5315–5329, Dec. 2015.

[28] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. A. De Guise, "Liver segmentation on CT and MR using Laplacian mesh optimization," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2110–2121, Sep. 2017.

[29] J. Cheng *et al.*, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1019–1032, Jun. 2013.

[30] A. Vial *et al.*, "The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: A review," *Transl. Cancer Res.*, vol. 7, no. 3, pp. 803–816, 2018.

[31] R. Li, L. Xing, S. Napel, and D. L. Rubin, *Radiomics and Radiogenomics: Technical Basis and Clinical Applications*. Boca Raton, FL, USA: CRC Press, 2019.

[32] Q. Luo *et al.*, "Segmentation of abdomen MR images using kernel graph cuts with shape priors," *Biomed. Eng. OnLine*, vol. 12, no. 1, p. 124, 2013.

[33] W. Wu, Z. Zhou, S. Wu, and Y. Zhang, "Automatic liver segmentation on volumetric CT images using supervoxel-based graph cuts," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–14, Mar. 2016.

[34] 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," *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.

[35] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2016, pp. 424–432.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*. [Online]. Available: https://arxiv.org/abs/1505.04597

[37] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[38] P. Hu, F. Wu, J. Peng, P. Liang, and D. Kong, "Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution," *Phys. Med. Biol.*, vol. 61, no. 24, p. 8676, 2016.

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.

[39] W. Qin *et al.*, "Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation," *Phys. Med. Biol.*, vol. 63, no. 9, May 2018, Art. no. 095017.

[40] B. Ibragimov and L. Xing, "Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks," *Med. Phys.*, vol. 44, no. 2, pp. 547–557, Feb. 2017.

[41] S. Nikolov *et al.*, "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy," 2018, *arXiv:1809.04430*. [Online]. Available: http://arxiv.org/abs/1809.04430

[42] Y. Guo, Y. Gao, and D. Shen, "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1077–1089, Apr. 2016.

[43] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, Mar. 2015.

[44] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 2015, pp. 234–241.

[46] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*. [Online]. Available: http://arxiv.org/abs/1608.06993

[47] M. Majurski *et al.*, "Cell image segmentation using generative adversarial networks, transfer learning, and augmentations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2019/html/CVMI/Majurski_Cell_Image_Segmentation_Using_Generative_Adversarial_Networks_Transfer_Learning_and_CVPRW_2019_paper.html

[48] Y. Wu and K. He, "Group normalization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[49] S. Shalev-Shwartz and A. Tewari, "Stochastic methods for ℓ_1-regularized loss minimization," *J. Mach. Learn. Res.*, vol. 12, pp. 1865–1892, Jun. 2011.

[50] A. Y. Ng, "Feature selection, L_1 vs. L_2 regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 78.

[51] G. Kang, J. Li, and D. Tao, "Shakeout: A new approach to regularized deep neural network training," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1245–1258, May 2018.

[52] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288,' Jan. 1996.

[53] Y. Shi, M. Lei, R. Ma, and L. Niu, "Learning robust auto-encoders with regularizer for linearity and sparsity," *IEEE Access*, vol. 7, pp. 17195–17206, 2019.

[54] M. Tofighi, T. Guo, J. K. P. Vanamala, and V. Monga, "Prior information guided regularized deep learning for cell nucleus detection," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2047–2058, Sep. 2019.

[55] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," 2018, *arXiv:1810.11654*. [Online]. Available: http://arxiv.org/abs/1810.11654

[56] S. Ruder, "An overview of multi-task learning in deep neural networks," 2017, *arXiv:1706.05098*. [Online]. Available: http://arxiv.org/abs/1706.05098

[57] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[58] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[60] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[61] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467

[62] A. L. Simpson *et al.*, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," 2019, *arXiv:1902.09063*. [Online]. Available: http://arxiv.org/abs/1902.09063

[63] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 2483–2493.

[64] X. Lian and J. Liu, "Revisit batch normalization: New understanding and refinement via composition optimization," in *Proc. 22nd Int. Conf. Artif. Intell. Statist.*, 2019, pp. 3254–3263.

[65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*. [Online]. Available: http://arxiv.org/abs/1512.03385

Authorized licensed use limited to: Hong Kong University of Science and Technology. Downloaded on March 16,2022 at 14:12:40 UTC from IEEE Xplore. Restrictions apply.