



# Thyroid nodule segmentation and classification in ultrasound images through intra- and inter-task consistent learning

Qingbo Kang<sup>a,b,c</sup>, Qicheng Lao<sup>a,b,c,e,\*</sup>, Yiyue Li<sup>a,b,c</sup>, Zekun Jiang<sup>a,b,c</sup>, Yue Qiu<sup>a,b,c</sup>,  
Shaoting Zhang<sup>c,e,f</sup>, Kang Li<sup>a,b,c,d,e,\*\*</sup>

<sup>a</sup> West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China

<sup>b</sup> Med-X Center for Informatics, Sichuan University, Chengdu, Sichuan 610041, China

<sup>c</sup> West China Hospital-SenseTime Joint Lab, Chengdu, Sichuan 610041, China

<sup>d</sup> Sichuan University - Pittsburgh Institute, Sichuan University, Chengdu, Sichuan 610207, China

<sup>e</sup> Shanghai Artificial Intelligence Laboratory, Shanghai 200030, China

<sup>f</sup> SenseTime Research, Shanghai 200233, China

## ARTICLE INFO

### Article history:

Received 6 November 2021

Revised 17 February 2022

Accepted 1 April 2022

Available online 25 April 2022

### Keywords:

Thyroid nodule

Ultrasound image

Segmentation and classification

Multi-task learning

Multi-stage learning

Task consistency

## ABSTRACT

Thyroid nodule segmentation and classification in ultrasound images are two essential but challenging tasks for computer-aided diagnosis of thyroid nodules. Since these two tasks are inherently related to each other and sharing some common features, solving them jointly with multi-task learning is a promising direction. However, both previous studies and our experimental results confirm the problem of inconsistent predictions among these related tasks. In this paper, we summarize two types of task inconsistency according to the relationship among different tasks: intra-task inconsistency between homogeneous tasks (e.g., both tasks are pixel-wise segmentation tasks); and inter-task inconsistency between heterogeneous tasks (e.g., pixel-wise segmentation task and categorical classification task). To address the task inconsistency problems, we propose intra- and inter-task consistent learning on top of the designed multi-stage and multi-task learning network to enforce the network learn consistent predictions for all the tasks during network training. Our experimental results based on a large clinical thyroid ultrasound image dataset indicate that the proposed intra- and inter-task consistent learning can effectively eliminate both types of task inconsistency and thus improve the performance of all tasks for thyroid nodule segmentation and classification.

© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

As one of the most common nodular lesions and endocrine carcinoma, thyroid nodule has been frequently diagnosed in adult population, with the prevalence of about 19% to 68% in clinical practice (Haugen et al., 2016). On the other hand, according to the statistics of 2018 global cancer worldwide, thyroid cancer ranked ninth in incidence and sixth in mortality (Bray et al., 2018). Although the majority of nodules are benign (noncancerous), there is a small percentage of them contains thyroid cancer, which is still curable if early diagnosed. Therefore accurate differentiation between benign and malignant thyroid nodules through non-invasive

methods can not only reduce potential patient cancer risk, but also avoid unnecessary fine-needle aspiration (FNA) and/or surgery.

Being a real-time, convenient, inexpensive and non-invasive imaging method, ultrasound (US) technique becomes a widely utilized tool for thyroid nodule diagnosis. In clinical examinations, radiologists usually identify the benign or malignant thyroid nodules by observing some important sonographic characteristics such as composition, echogenicity, shape and margin properties (Tessler et al., 2017). However, due to the relatively low quality, resolution and contrast, as well as speckle noises and echo perturbations, US based thyroid nodule assessment is heavily dependent on the clinical experiences of radiologists, and thus the diagnosis results are subjective.

In order to tackle this problem, many computer-aided diagnosis (CAD) systems for thyroid nodule diagnosis in US images have been proposed (Koundal et al., 2012; Chen et al., 2020). Nodule segmentation and classification are two key and basic tasks in CAD systems. Since these two tasks are highly related to each other

\* Corresponding author at: West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China.

\*\* Corresponding author at: Shanghai Artificial Intelligence Laboratory, Shanghai 200030, China.

E-mail addresses: [qicheng.lao@gmail.com](mailto:qicheng.lao@gmail.com) (Q. Lao), [likang@wchscu.cn](mailto:likang@wchscu.cn) (K. Li).

and sharing some common image features (e.g. nodule boundary characteristics can be used as a clue both for classification and segmentation), solving these two tasks jointly in one unified model is a promising direction. However, over the past decades, previous studies on thyroid nodule diagnosis in US images often treated the two tasks separately.

Multi-task learning (MTL) is a learning paradigm which aims at learning multiple related tasks in parallel, and improves generalization abilities of all tasks by sharing learned representations across numerous training signals. MTL have been employed in many medical image analysis tasks (Xie et al., 2018; Wang et al., 2018; Ployout et al., 2018; Qu et al., 2019; Singh et al., 2019; He et al., 2020, 2021; Zhou et al., 2021). Recently, a similar study that focuses on the segmentation and classification of tumors in 3D automated breast ultrasound (ABUS) images has been presented in Zhou et al. (2021). In their network, a classification branch is added on the bottom of encoder-decoder style segmentation network thus enables the whole model learn the two related tasks simultaneously. It has been proven that learning nodule classification and segmentation jointly can boost the performance of both tasks. In fact, the proposed MTL network in Zhou et al. (2021) is functionally equivalent to a multi-class segmentation network which also outputs segmentation maps and categorical predictions. Yet, it is not known whether the addition of one classification branch on the bottom of the encoder in the multi-class segmentation network could further improve the performance of each task.

To answer this question, in this study, we focus on applying MTL network which performs multi-class segmentation and classification on thyroid nodules in US images. Moreover, given a MTL network that learns multi-class segmentation and classification in parallel, one natural assumption is that a robust model should learn consistent predictions between the two tasks since these two tasks are inherently related to each other. However, previous studies have shown that this assumption is not always tenable in many domains including medical imaging (Luo et al., 2021; Seo et al., 2021; He et al., 2021). According to the relationship among different tasks, we summarize the task inconsistency into two different types: intra-task inconsistency and inter-task inconsistency. Specifically, we refer the inconsistency between homogeneous tasks as intra-task inconsistency, e.g., if both tasks are pixel-wise tasks (also known as high-dimensional tasks, such as semantic segmentation). The inter-task inconsistency, on the other hand, is the inconsistency between heterogeneous tasks, e.g., between a pixel-wise task and categorical tasks (also known as low-dimensional tasks, such as image classification). We refer the effort of trying to tackle the inconsistency among multiple related tasks as task consistent learning. Various work on task consistent learning has been proposed, for example, Zamir et al. (2020) designed a inference-path invariance across multiple pixel-wise tasks to solve the task inconsistent problem for natural images. However, as they explained, their work is limited to intra-task consistency. The same limitation also exists in other task consistency related works (Lu et al., 2021; Luo et al., 2021; Seo et al., 2021; He et al., 2021). To the extent of our knowledge, there is few work conducting research on the inter-task consistency, since compared with the intra-task consistency, the inter-task consistency is much more challenging to deal with as it relies on the communications between heterogeneous tasks whose output predictions may have completely different meanings and (or) shapes. Therefore, it requires extra engineering or algorithms to enforce multiple predictions from heterogeneous tasks to be consistent with each other.

In this paper, unlike the aforementioned existing work which only focuses on intra-task consistency, we establish a new paradigm of task consistent learning where we introduce inter-task consistency to complement with intra-task consistency for task consistent learning. The challenge of inter-task consistent learning

is to enforce different forms of predictions from related inter-tasks coordinate with each other and synergically improve the overall performance. To do so, we develop a novel algorithm in this work for computing the inter-task inconsistency between thyroid nodule segmentation and classification, and by minimizing such inconsistency we achieve task consistent learning. In addition, to better leverage the effectiveness of task consistency, we design a multi-stage multi-task learning (MS-MTL) network for thyroid nodule diagnosis, which performs three tasks in two stages, i.e., binary segmentation and classification in the first stage, and multi-class segmentation in the second stage. Based on the inherent relationship among the three tasks, we propose three instances of task consistent learning: one for intra-tasks and the other two for inter-tasks. More concretely, since the multi-class segmentation and the binary segmentation are naturally similar to each other, we enforce the intra-task consistency between them during the network training. On the other hand, because the classification is a categorical task, which is inhomogeneous to the two segmentation tasks, we design two kinds of inter-task consistency: one is only between the classification and multi-class segmentation; and the other is for all the three tasks. Finally, we conduct extensive experiments on a large clinical thyroid US dataset to evaluate the effectiveness of our proposed intra- and inter-task consistent learning with the MS-MTL network for thyroid nodule segmentation and classification. The main contributions of our paper are listed as follows:

- For the first time, we apply multi-task learning on thyroid nodule diagnosis where we design a multi-stage and multi-task learning framework for the joint classification and multi-class segmentation of thyroid nodules in ultrasound images. It performs three tasks in two stages simultaneously: binary segmentation and classification in the first stage, and multi-class segmentation in the second stage.
- We formulate a new paradigm of task consistency by emphasizing on both intra- and inter-task consistency to enforce the MS-MTL network learn consistent predictions for all the tasks. Specifically, our task consistency consists of three parts and can be categorized into two categories. The intra-task consistency means the consistency between homogeneous tasks, while the inter-task consistency means the consistency between heterogeneous tasks.
- We evaluate the effectiveness of our proposed intra- and inter-task consistent learning on a large clinical thyroid ultrasound image dataset which is collected in West China Hospital. The experimental results show that the proposed intra- and inter-task consistent learning can effectively eliminate both types of task inconsistency and thus improve the performance of all tasks for thyroid nodule segmentation and classification.

## 2. Related work

### 2.1. Ultrasound thyroid nodule segmentation and classification

Over the past decades, ultrasound thyroid nodule segmentation approaches based on traditional image processing techniques can be roughly categorized into four types: shape and contour based (Maroulis et al., 2007; Nugroho et al., 2015; Du and Sang, 2015; Koundal et al., 2016), region based (Zhao et al., 2013; Alrubaidi et al., 2016), machine learning based (Chang et al., 2009; Keramidas et al., 2012) and hybrid methods (Legakis et al., 2011; Zhou, 2016). More detailed review is presented in Chen et al. (2020). Research on thyroid nodule classification in US images is mainly based on the extraction of hand-designed features such as sonographic textural features (Chen et al., 2010; Katsigiannis et al., 2010) or Local Binary Patterns (LBP) (Iakovidis et al., 2008), and then in combination with some machine learning classifiers such

as Support Vector Machine (SVM) (Chang et al., 2010), Linear Discriminant Analysis (LDA) (Luo et al., 2011) to classify the nodule as benign or malignant. Although these approaches can achieve satisfactory segmentation results or classification accuracy, in some scenarios, their performance is heavily depends on the selection of manually designed features and therefore the generalization ability of these approaches is limited.

Recently, many approaches based on Convolutional Neural Network (CNN) have been proposed for US thyroid nodule segmentation (Ma et al., 2017a; Ying et al., 2018; Ding et al., 2019; Ouahabi and Taleb-Ahmed, 2021) and classification (Ma et al., 2017b; Liu et al., 2017; Chi et al., 2017). For the segmentation problem, Ma et al. (2017a) converted the nodule segmentation task into a patch-based classification task, then a CNN model with 15 convolutional layers is utilized to solve the patch classification problem. Ying et al. (2018) designed a two-step framework for nodule segmentation, i.e., the ROI of nodule is extracted from the input US image in the first step, after which a FCN was used to perform pixel-wise nodule segmentation in the second step. Ding et al. (2019) improved the original U-Net with a modified residual unit and attention mechanism to develop a segmentation approach. Ouahabi and Taleb-Ahmed (2021) proposed an efficient real-time segmentation method on the basis of U-Net combined with dense connectivity, factorized convolution and atrous convolution. For the classification problem, Ma et al. (2017b) developed a hybrid classification method which is a fusion of features from two different trained CNN models. Liu et al. (2017) combined CNN with traditional image features such as Histogram of Oriented Gradient (HOG) to improve the classification accuracy. Chi et al. (2017) utilized GoogLeNet (Szegedy et al., 2015) to extract thyroid US image features and then a Random Forest classifier was used to classify the nodule as benign or malignant. In this work, we apply multi-task learning for the joint classification and multi-class segmentation of thyroid nodules in ultrasound images.

## 2.2. Multi-task learning

MTL aims to learn multiple different but related tasks in parallel with one unified model. The principle objective of such learning paradigm is to boost the generalization abilities by sharing learned representations for all tasks. Zhang and Yang (2021) provided a comprehensive review on MTL. Up to now, MTL has been widely used in many domains including medical imaging especially for image classification and segmentation (Xie et al., 2018; Wang et al., 2018; Playout et al., 2018; Qu et al., 2019; Singh et al., 2019; He et al., 2020; Zhou et al., 2021; He et al., 2021). Xie et al. (2018) introduced a two-stage approach for breast US image classification and segmentation. In the first stage, a pre-trained ResNet (He et al., 2016) was utilized to classify input image as normal or cancerous, then in the second stage, an improved Mask R-CNN (He et al., 2017) was adopted to segment tumors from the classified cancerous images. Qu et al. (2019) connected a prediction network which performs a multi-class segmentation with a perceptual loss network used for segmentation refinement to jointly segment and classify different types of nuclei (tumor, lymphocyte and stroma nuclei) in histopathology images. He et al. (2020) designed a MTL framework for organs at risk segmentation in CT images. Specifically, a classification head was added in parallel with the segmentation head for encoder-decoder style networks, which performs a multi-label classification task that indicates whether the corresponding organ is existed in the segmentation mask. Wang et al. (2018) presented a MTL CNN model to segment and classify bone surfaces simultaneously in US images. The CNN model is a modified U-Net by adding a classification branch after the encoder. Following this,

Zhou et al. (2021) also added a light-weighted multi-scale classification network to the bottom of V-Net (Milletari et al., 2016) to construct a MTL framework for joint tumor segmentation and classification in 3D ABUS images.

## 2.3. Multi-stage learning

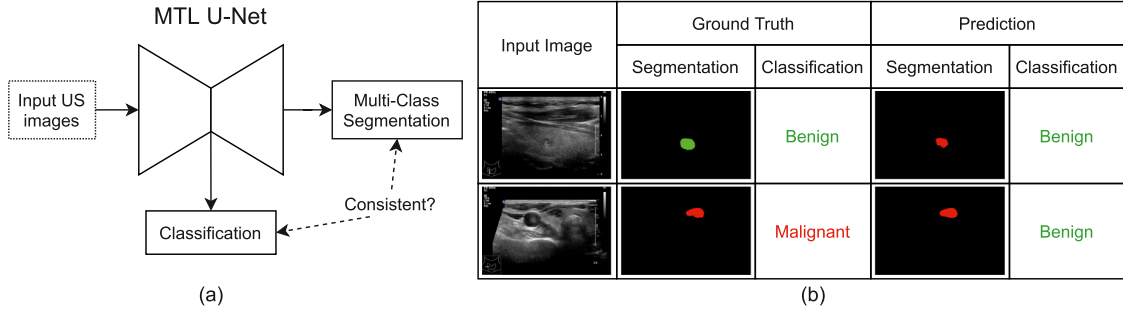
The purpose of MSL is to decompose the complex learning process of one task into multiple intermediate stages or steps, therefore decrease the learning difficulty and smooth the learning curve of the final task. Several approaches based on MSL for medical image analysis have been proposed. Takahama et al. (2019) presented a multi-stage classification framework for tumor/normal predication on whole-slide histopathological images, where the features are firstly extracted from local image patches through a classification model in the first stage, and then followed by a segmentation model to obtain the tumor/normal predication result. Bi et al. (2017) developed a multi-stage segmentation method based on stacking multiple Fully Convolutional Networks (FCNs) for skin lesion segmentation, in which the FCNs in early-stage capture rough semantic features while the FCNs in late-stage learn the fine-grained subtle features. Kang et al. (2019) designed a two-stage learning approach for nuclei segmentation in histopathological images, where a nuclei-boundary prediction task was introduced in the first stage to address the segmentation of touching nuclei. Chen et al. (2021) proposed a multi-stage framework based on two stacked CNN models for aortic dissection segmentation, where the first CNN in the first stage aims to segment the aortic trunk and all branches, and then the second CNN in the second stage was responsible for separating true lumen and false lumen on the basis of the segmented aortic trunk in the first stage.

## 2.4. Task consistency

In computer vision or medical imaging field, various kinds of task consistency have been studied, including but not limited to cycle consistency, data consistency and task consistency. Cycle consistency (Zhu et al., 2017) has been extensively used in image-to-image translations between different domains (Dwivedi et al., 2019), whereas data consistency refers to that one or multiple models should have consistent predictions for the same data sample with different transformations or perturbations and has been broadly used in semi-supervised learning (Bortsova et al., 2019; Li et al., 2020; Mittal et al., 2019; Ouali et al., 2020; Sajjadi et al., 2016; Tarvainen, Valpola). In contrast, task consistency means the predictions across multiple tasks of one model should be consistent under the condition that these tasks are intrinsically related to each other (Zou et al., 2018; Zamir et al., 2020; Lu et al., 2021; Luo et al., 2021; Seo et al., 2021; He et al., 2021). In this work, we focus on the task consistency.

For computer vision tasks, Zou et al. (2018) designed an unsupervised learning approach based on the consistency between depth prediction and optical flow estimation tasks. Zamir et al. (2020) formulated the task consistency as inference-path invariance across a graph of different tasks and evaluated on high-dimensional/pixel-wise tasks (e.g., surface normal predication). Lu et al. (2021) utilized task consistency to train a collective of pixel-wise tasks including depth prediction, ego-motion and semantic segmentation. The main drawback of these works is that the consistency on low-dimensional/categorical tasks such as classification has not been exploited.

In medical image analysis field, Luo et al. (2021) proposed a dual-task consistency method and applied it on semi-supervised medical image segmentation. The consistency between a level set representation prediction and an ordinary segmentation prediction



**Fig. 1.** Motivation of our proposed approach. (a) Schematic diagram of the MTL U-Net. (b) Inconsistent predictions between multi-class segmentation and classification. Benign nodule is shown in green and malignant nodule is shown in red.

was introduced to regularize the training of MTL model. However, only intra-task consistency (i.e., segmentation) was considered in this work. The same limitation is also occurred in (Seo et al., 2021; He et al., 2021). In He et al. (2021), a hierarchically-fused U-Net (HF-UNet) was proposed to solve prostate segmentation in CT images by utilizing MTL. Specifically, an auxiliary prostate boundary regression task was added into their MTL framework, where one U-Net with two branches was designed to solve the two tasks and attention-based task consistency learning blocks were used to communicate interactive information between these two branches. Although they claimed inter-task relevance was used in their network, since the two tasks (boundary regression and prostate segmentation) are homogeneous (i.e., both are pixel-wise tasks), we argue that their task consistency is still intra-task consistency. To the extent of our knowledge, our proposed approach is the first to introduce inter-task consistency and combine it with intra-task consistency for MTL.

### 3. Method

#### 3.1. Motivation

In previous MTL research for joint segmentation and classification (Wang et al., 2018; Zhou et al., 2021), the MTL network usually performs two tasks simultaneously: one segmentation and one classification, where the segmentation aims to segment binary pixel-wise target contours while the classification intends to classify different types of the target (such as benign or malignant for tumor). It has been proven that learning these two tasks jointly can boost both the segmentation and classification performance (Zhou et al., 2021). In fact, such MTL network is functionally equivalent to a segmentation network that performs multi-class segmentation. However, it is not known whether the addition of a classification branch can further boost the performance of multi-class segmentation. In addition, since these two tasks are inherently related to each other, another crucial question is that whether the predictions from these two tasks are consistent with each other.

In order to answer these questions, we construct a MTL U-Net for thyroid nodule segmentation and classification by inserting a classification branch after the encoder of the U-Net. As shown in the schematic diagram of the MTL U-Net (Fig. 1(a)), it performs two tasks: a 3-class segmentation task (i.e., background, benign or malignant thyroid nodule) and a classification task (i.e., benign or malignant nodule). Since these two tasks are inherently related to each other given the same input, one natural assumption is that the predictions should be consistent with each other. However, we observed in our experiments that this assumption is not always tenable. For example, the first row in Fig. 1(b) shows an example case where the classification is correctly predicted while the

segmentation prediction is wrong, i.e., the 3-class segmentation map outputs a benign nodule (shown in green color) as malignant (shown in red color). Note that although the most parts of the nodule have been segmented correctly, we still consider the segmentation is wrong because the predicted class of the segmentation nodule is wrong. Inversely, another example shown in the second row in Fig. 1(b) gives correct segmentation prediction but wrong classification prediction. Statistically, in a test set of 897 thyroid ultrasound images, 130 images (14.49%) have inconsistent predictions, among which, 67% of them have correct classification predictions but wrong segmentation predictions, while 18% of them have wrong classification predictions but correct segmentation predictions. The rest (15%) have wrong predictions on both classification and segmentation.

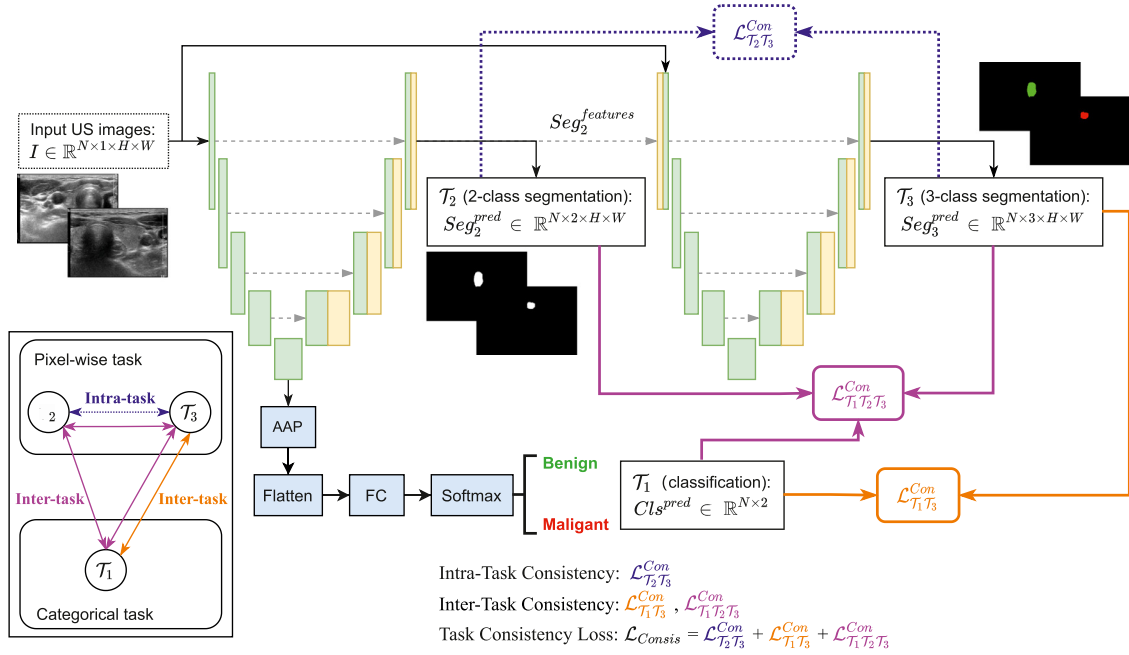
Motivated by the observed inconsistency phenomenon when applying MTL to thyroid nodule segmentation and classification in US images, we propose a new paradigm of intra- and inter-task consistent learning to enforce the MTL network learn consistent predictions among different tasks during network training. In order to better leverage the effectiveness of task consistency, we design a multi-stage and multi-task learning network where different stages in the network can output different tasks of interest. Depending on the relationship among the tasks, the task consistency is introduced accordingly in the network to synergically improve the performance of all tasks.

#### 3.2. Multi-stage multi-task learning network

The MS-MTL network is illustrated in Fig. 2. One MTL U-Net is used in the first stage, which performs two tasks and generates two predictions: thyroid nodule binary segmentation (background vs. nodule) and thyroid nodule classification (benign vs. malignant). Specifically, the classification branch is added after the encoder of U-Net, which consists of one adaptive average pooling (AAP) layer for reducing the dimension of the output features extracted from the encoder to  $1 \times 1$ , one flatten layer, one fully connected (FC) layer and one softmax layer. The decoder of the MTL U-Net performs conventional binary segmentation task. Both of the two tasks share the same encoder of the MTL U-Net. After the MTL U-Net in the first stage, another U-Net is utilized in the second stage to perform a 3-class segmentation task (background, benign or malignant nodule). The input of the second U-Net is the concatenation of the features before the softmax layer in the first MTL U-Net and the original input image.

Given a batch of input US images  $I \in \mathbb{R}^{N \times 1 \times H \times W}$  where  $N$  denotes the batch size, and  $H$  and  $W$  represent image height and width. The corresponding one-hot encoding ground truth of 2-class segmentation, 3-class segmentation and classification are denoted by  $Seg_2^{GT}$ ,  $Seg_3^{GT}$ ,  $Cls^{GT}$ , respectively. We can then define the following:





**Fig. 2.** Overview architecture of the MS-MTL network with our proposed task consistency. Since  $T_2$  and  $T_3$  are both pixel-wise segmentation task, the consistency loss between them ( $\mathcal{L}_{T_2T_3}^{Con}$ ) is intra-task consistency. While  $T_1$  is a categorical classification task, therefore any consistency loss related to  $T_1$  ( $\mathcal{L}_{T_1T_3}^{Con}$  or  $\mathcal{L}_{T_1T_2}^{Con}$ ) belongs to inter-task consistency. The task consistency loss is the summation of intra- and inter-task consistency losses.

1. The 2-class segmentation prediction probability maps as  $Seg_2^{pred} \in \mathbb{R}^{N \times 2 \times H \times W}$ , in which the 1-st map is background and the 2-nd is nodule;
2. The 3-class segmentation prediction probability maps as  $Seg_3^{pred} \in \mathbb{R}^{N \times 3 \times H \times W}$ , in which the 1-st map is background, the 2-nd is benign nodule and the 3-rd is malignant nodule;
3. The classification prediction probabilities as  $Cls^{pred} \in \mathbb{R}^{N \times 2}$ , where the 1-st element means benign and the 2-nd means malignant.

With the above definitions, the two tasks in the first stage ( $T_1$ ,  $T_2$ ) and the task in the second stage ( $T_3$ ) can be written as:

$$\begin{aligned} T_1 &= I \mapsto Cls^{pred} \\ T_2 &= I \mapsto Seg_2^{pred} \\ T_3 &= \{I, Seg_2^{features}\} \mapsto Seg_3^{pred}, \end{aligned} \quad (1)$$

where  $Seg_2^{features}$  denotes the features before softmax layer from the MTL U-Net in the first stage.

For the segmentation tasks ( $T_2$  and  $T_3$ ), we use the dice loss:

$$\mathcal{L}_{DICE}(Seg^{GT}, Seg^{pred}) = 1 - \frac{2 \sum_{n=1}^N (Seg_{(n)}^{GT} Seg_{(n)}^{pred}) + 1}{\sum_{n=1}^N Seg_{(n)}^{GT} + \sum_{n=1}^N Seg_{(n)}^{pred} + 1}, \quad (2)$$

where  $Seg_{(n)}^{GT}$  represents the  $n$ th segmentation map ground truth and  $Seg_{(n)}^{pred}$  denotes the  $n$ th predicted segmentation probabilities. For the classification task ( $T_1$ ), we use binary cross entropy (BCE) loss:

$$\mathcal{L}_{BCE}(Cls^{GT}, Cls^{pred}) = -\frac{1}{N} \sum_{n=1}^N \left[ Cls_{(n)}^{GT} \log(Cls_{(n)}^{pred}) + (1 - Cls_{(n)}^{GT}) \log(1 - Cls_{(n)}^{pred}) \right], \quad (3)$$

where  $Cls_{(n)}^{GT}$  and  $Cls_{(n)}^{pred}$  means the ground truth label (i.e., 0 for benign and 1 for malignant nodule) and predicted probability for the  $n$ th input US image.

### 3.3. Intra- and inter-task consistent learning

Here, we introduce the proposed intra- and inter-task consistent learning for addressing the inconsistent problems when using the above MS-MTL network. As mentioned before, the task consistency involves two different types of consistency: intra-task and inter-task consistency. We first present the trivial solution for dealing with the intra-task consistency for homogeneous tasks (Section 3.3.1). We then focus on the more challenging part, i.e., the inter-task consistency for inhomogeneous tasks (Section 3.3.2), which to the best of our knowledge has not been studied before.

#### 3.3.1. Intra-task consistency

The intra-task consistency in this work contains the consistency between the 2-class segmentation and 3-class segmentation tasks ( $T_2$  and  $T_3$ ) since both of them are homogeneous pixel-wise tasks. The process for computing this consistency is relatively straightforward. To do so, we first transform the shape of 3-class segmentation prediction to the shape of 2-class segmentation prediction, and then the dice loss can be used to compute the consistency loss.

Specifically, given their corresponding predicted segmentation probabilities  $Seg_2^{pred}$  and  $Seg_3^{pred}$ , we first reshape  $Seg_3^{pred} \in \mathbb{R}^{N \times 3 \times H \times W}$  to get  $Seg_3 - as - Seg_2 \in \mathbb{R}^{N \times 2 \times H \times W}$  by merging the benign and malignant maps (the 2-nd and 3-rd maps) of  $Seg_3^{pred}$  while keeping the background map (the 1-st map) unchanged, i.e.,  $Seg_3 - as - Seg_2[:, :, 1, :] = Seg_3^{pred}[:, :, 1, :] + Seg_3^{pred}[:, :, 2, :]$  and  $Seg_3 - as - Seg_2[:, :, 0, :] = Seg_3^{pred}[:, :, 0, :]$ . After the reshaping step, both of the two tensors  $Seg_3 - as - Seg_2$  and  $Seg_2^{pred}$  have the same shape. Finally, we apply dice loss on the two tensors to compute the consistency loss:

$$\mathcal{L}_{T_2T_3}^{Con} = \mathcal{L}_{DICE}(Seg_3 - as - Seg_2, Seg_2^{pred}). \quad (4)$$

#### 3.3.2. Inter-task consistency

The inter-task consistency refers to any consistency that involves inhomogeneous tasks. In this work, we propose to use two

kinds of inter-task consistency losses: the consistency between the classification task and 3-class segmentation task:  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$ , and the consistency among three tasks:  $\mathcal{L}_{T_1 T_2 T_3}^{\text{Con}}$ .

The philosophy of computing  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$  is to transform the 3-class segmentation prediction results to classification predictions. Concretely, the shape of  $\text{Seg}_3^{\text{pred}} \in \mathbb{R}^{N \times 3 \times H \times W}$  needs to be converted to the same shape as  $\text{Cls}^{\text{pred}} \in \mathbb{R}^{N \times 2}$ . The procedure of computing  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$  is shown in Algorithm 1. In general, we convert the 3-class

**Algorithm 1** The procedure of computing consistency loss between classification and 3-class segmentation tasks.

**Input:** 3-class segmentation prediction probability maps,  $\text{Seg}_3^{\text{pred}} \in \mathbb{R}^{N \times 3 \times H \times W}$ , where  $N$  denotes the batch size,  $H$  and  $W$  stand for the height and width of the segmentation mask image, respectively.

**Input:** Classification prediction probabilities,  $\text{Cls}^{\text{pred}} \in \mathbb{R}^{N \times 2}$ .

**Output:** Consistency loss between classification and 3-class segmentation tasks:  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$ .

- 1: Create a Boolean tensor  $\text{SegMaxIndicator} \in \mathbb{R}^{N \times 3 \times H \times W}$  which has exactly the same shape as  $\text{Seg}_3$  and indicates whether the corresponding element in the  $\text{Seg}_3$  is the maximum along the 2-nd channel dimension (the class channel):  $\text{SegMaxMask} \leftarrow \max_{\text{dim}=2}(\text{Seg}_3^{\text{pred}}) \leq \text{Seg}_3^{\text{pred}}$
- 2: Calculate  $\text{SegMax} \in \mathbb{R}^{N \times 3 \times H \times W}$  by only keeping the maximum probability in  $\text{Seg}_3^{\text{pred}}$  and setting others to 0:  $\text{SegMax} \leftarrow \text{Seg}_3^{\text{pred}} \times \text{SegMaxMask}$
- 3: Transform  $\text{SegMax} \in \mathbb{R}^{N \times 3 \times H \times W}$  to  $\text{SegMaxMean} \in \mathbb{R}^{N \times 3}$  by computing the average value of the maximums in  $\text{SegMax}$  along the last two dimensions (height and width dimensions):  $\text{SegMaxMean} \leftarrow \frac{\sum_{i=1}^H \sum_{j=1}^W \text{SegMax}(:, :, i, j)}{\sum_{i=1}^H \sum_{j=1}^W \text{SegMaxMask}(:, :, i, j) + \epsilon}$  where  $\epsilon$  is used for avoiding division by 0 and is set to  $1e-3$ .
- 4: Convert  $\text{SegMaxMean} \in \mathbb{R}^{N \times 3}$  to  $\text{Seg} - \text{as} - \text{Cls} \in \mathbb{R}^{N \times 2}$  via applying softmax function to the last two elements of  $\text{SegMaxMean}$  (removing the 1st background element):  $\text{Seg} - \text{as} - \text{Cls} \leftarrow \text{Softmax}(\text{SegMaxMean}[:, 2:])$
- 5: Apply soft cross entropy loss function  $\mathcal{L}_{\text{SCE}}$  to  $\text{Seg} - \text{as} - \text{Cls} \in \mathbb{R}^{N \times 2}$  and  $\text{Cls}^{\text{pred}} \in \mathbb{R}^{N \times 2}$ . Since SCE function is order sensitive, the final loss is the summation of two different orders:  $\mathcal{L}_{T_1 T_3}^{\text{Con}} \leftarrow \mathcal{L}_{\text{SCE}}(\text{Seg} - \text{as} - \text{Cls}, \text{Cls}^{\text{pred}}) + \mathcal{L}_{\text{SCE}}(\text{Cls}^{\text{pred}}, \text{Seg} - \text{as} - \text{Cls})$
- 6: **return**  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$

pixel-wise segmentation predictions into the 2 single probabilities which sums to 1. Specifically, we firstly obtain  $\text{SegMaxMean} \in \mathbb{R}^{N \times 3}$  by averaging segmentation predictions on all ‘max-probability’ locations along the last two dimensions (height and width). Here, ‘max-probability’ location means the probability at the corresponding location is the maximum along the 2-nd channel dimension (class). We will show later in our ablation study that compared to using ‘all-probability’ locations, selecting ‘max-probability’ locations is critical for a meaningful classification prediction, where the noises from background pixels can be filtered out. Secondly, we obtain  $\text{Seg} - \text{as} - \text{Cls} \in \mathbb{R}^{N \times 2}$  by removing the first element of  $\text{SegMaxMean}$  which corresponds to the background class and applying softmax function to the remaining two elements. The reason behind the softmax operation is that we need these two elements which denote the probabilities of benign or malignant sum to 1. Finally, since both  $\text{Seg} - \text{as} - \text{Cls}$  and  $\text{Cls}^{\text{pred}}$  are soft labels and with the same shape, soft cross entropy (SCE) loss function  $\mathcal{L}_{\text{SCE}}$  is used to measure the loss between them. The SCE is defined as

follows:

$$\mathcal{L}_{\text{SCE}}(Y^{GT}, Y^{\text{pred}}) = -\frac{1}{N} \sum_{j=1}^N \sum_{i=1}^K Y_{ji}^{GT} \log(Y_{ji}^{\text{pred}}), \quad (5)$$

where  $K$  indicates the number of classes and  $Y_{ji}^{GT}$  denotes the ground truth probability of  $j$ th image belonging to  $i$ th class while  $Y_{ji}^{\text{pred}}$  stands for the predicted probability of  $j$ th image belonging to  $i$ th class.

We use summation of two SCE losses with different orders which may have different meaning. Specifically,  $\mathcal{L}_{\text{SCE}}(\text{Seg} - \text{as} - \text{Cls}, \text{Cls}^{\text{pred}})$  treats the segmentation results as ground truth and enforces the classification learn consistent results with segmentation, while  $\mathcal{L}_{\text{SCE}}(\text{Cls}^{\text{pred}}, \text{Seg} - \text{as} - \text{Cls})$  treats the classification results as ground truth and encourages the segmentation to be consistent with the classification. We will show later in the experiments (Section 5.3) that these two different orders are complement to each other. Therefore,  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$  can be calculated by the following:

$$\mathcal{L}_{T_1 T_3}^{\text{Con}} = \mathcal{L}_{\text{SCE}}(\text{Seg} - \text{as} - \text{Cls}, \text{Cls}^{\text{pred}}) + \mathcal{L}_{\text{SCE}}(\text{Cls}^{\text{pred}}, \text{Seg} - \text{as} - \text{Cls}). \quad (6)$$

To Facilitate the understanding, we give an illustration of the inter-task consistency between classification and 3-class segmentation ( $\mathcal{L}_{T_1 T_3}^{\text{Con}}$ ) in Fig. 3. As shown in the figure, two upper rows have consistent predictions while two bottom rows have inconsistent predictions. This difference can be effectively reflected by the corresponding computed inter-task consistency loss  $\mathcal{L}_{T_1 T_3}^{\text{Con}}$ , i.e., the losses of the two bottom rows are much higher than those of the two upper rows. In addition, the  $\text{Seg} - \text{as} - \text{Cls}$  computed with ‘max-probability’ locations gives meaningful predictions whereas the predictions by ‘all-probability’ locations are unreliable random guesses.

Another kind of inter-task consistency loss we propose in this work is the consistency loss among all the three tasks (2-class segmentation, 3-class segmentation and classification) in our MS-MTL network, i.e.,  $\mathcal{L}_{T_1 T_2 T_3}^{\text{Con}}$ . The procedure of computing  $\mathcal{L}_{T_1 T_2 T_3}^{\text{Con}}$  is presented in Algorithm 2. Specifically, we first generate a 3-class segmentation map  $\text{Seg}_2 - \text{as} - \text{Seg}_3 \in \mathbb{R}^{N \times 3 \times H \times W}$  by incorporating the 2-class segmentation map  $\text{Seg}_2^{\text{pred}} \in \mathbb{R}^{N \times 2 \times H \times W}$  with the classification result  $\text{Cls}^{\text{pred}} \in \mathbb{R}^{N \times 2}$ , and then the dice loss  $\mathcal{L}_{\text{DICE}}$  is applied on the generated 3-class segmentation map  $\text{Seg}_2 - \text{as} - \text{Seg}_3$  and the predicted 3-class segmentation map  $\text{Seg}_3^{\text{pred}} \in \mathbb{R}^{N \times 3 \times H \times W}$  to compute the loss:

$$\mathcal{L}_{T_1 T_2 T_3}^{\text{Con}} = \mathcal{L}_{\text{DICE}}(\text{Seg}_2 - \text{as} - \text{Seg}_3, \text{Seg}_3^{\text{pred}}). \quad (7)$$

### 3.4. Total loss function




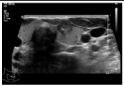


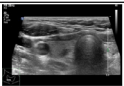

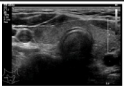


In summary, for the intra- and inter-task consistent learning, we have one intra-task consistency loss: the consistency between 2-class segmentation ( $T_2$ ) and 3-class segmentation ( $T_3$ ); and two inter-task consistency losses: the consistency between classification ( $T_1$ ) and 3-class segmentation ( $T_3$ ), and the consistency among all the three tasks ( $T_1$ ,  $T_2$ , and  $T_3$ ). Given the equations described above (Eqs. (4), (6) and (7)), the full task consistency loss  $\mathcal{L}_{\text{Consis}}$  is:

$$\mathcal{L}_{\text{Consis}} = \mathcal{L}_{T_2 T_3}^{\text{Con}} + \mathcal{L}_{T_1 T_3}^{\text{Con}} + \mathcal{L}_{T_1 T_2 T_3}^{\text{Con}}. \quad (8)$$

Since we use the dice loss for 2-class segmentation and 3-class segmentation, the BCE loss for classification, the total loss of our MS-MTL network with intra- and inter-task consistent learning is the weighted summation of these three individual task losses and the full task consistency loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}}^{T_1} + \mathcal{L}_{\text{DICE}}^{T_2} + \mathcal{L}_{\text{DICE}}^{T_3} + \lambda \mathcal{L}_{\text{Consis}}, \quad (9)$$

where  $\lambda$  is the weight used to regulate the contribution of task consistent learning. We set  $\lambda$  to 0.3 throughout this work. Note

Input US Image	$Seg_3^{GT}$	$Cls^{GT}$	$Seg_3^{pred}$	$Seg-as-Cls$ (all probability)	$Seg-as-Cls$ (max probability)	$Cls^{pred}$	$\mathcal{L}_{T_1 T_2 T_3}^{Con}$	
		[1, 0]		[0.502, 0.498]	[0.731, 0.269]	[1, 0]	1.0711	Consistent
		[0, 1]		[0.499, 0.501]	[0.269, 0.731]	[0, 1]	1.0716	
		[1, 0]		[0.500, 0.500]	[0.270, 0.730]	[0.999, 0.001]	1.9918	Inconsistent
		[0, 1]		[0.499, 0.501]	[0.270, 0.730]	[0.986, 0.014]	1.9738	

**Fig. 3.** An illustration of inter-task consistency  $\mathcal{L}_{T_1 T_2 T_3}^{Con}$ : the consistency between classification and 3-class segmentation. Benign nodules are shown in green and malignant nodules are shown in red.

**Algorithm 2** The procedure of computing consistency loss among all the three tasks (2-class segmentation, 3-class segmentation and classification) in our MS-MTL network.

**Input:** 2-class segmentation prediction probability maps,  $Seg_2^{pred} \in \mathbb{R}^{N \times 2 \times H \times W}$ .  
**Input:** 3-class segmentation prediction probability maps,  $Seg_3^{pred} \in \mathbb{R}^{N \times 3 \times H \times W}$ .  
**Input:** Classification prediction probabilities,  $Cls^{pred} \in \mathbb{R}^{N \times 2}$ .  
**Output:** Consistency loss among 2-class segmentation, 3-class segmentation and classification tasks:  $\mathcal{L}_{T_1 T_2 T_3}^{Con}$ .

- 1: Firstly, obtain classification label predictions  $Labels^{pred} \in \mathbb{R}^{N \times 1}$  by taking the maximum probability index of  $Cls^{pred} \in \mathbb{R}^{N \times 2}$ :  $Labels^{pred} \leftarrow \arg \max_{dim=2} (Cls^{pred})$
- 2: Secondly, incorporate the 2-class segmentation maps  $Seg_2^{pred}$  with the classification label predictions  $Labels^{pred}$  to generate a 3-class segmentation maps  $Seg_2 - as - Seg_3 \in \mathbb{R}^{N \times 3 \times H \times W}$ .
- 3: **for** each  $i \in [1, N]$  **do**
- 4:   **if**  $Labels^{pred}[i] = 0$  **then**
- 5:      $Seg_2 - as - Seg_3[i, 1, :, :] \leftarrow Seg_2^{pred}[i, :, :, :]$
- 6:   **else**
- 7:      $Seg_2 - as - Seg_3[i, 2, :, :] \leftarrow Seg_2^{pred}[i, :, :, :]$
- 8:   **end if**
- 9: **end for**
- 10: Lastly, apply DICE loss  $\mathcal{L}_{DICE}$  on the generated 3-class segmentation maps  $Seg_2 - as - Seg_3$  and the predicted 3-class segmentation maps  $Seg_3^{pred}$  to compute the final loss:  $\mathcal{L}_{T_1 T_2 T_3}^{Con} \leftarrow \mathcal{L}_{DICE}(Seg_2-as-Seg_3, Seg_3^{pred})$
- 11: **return**  $\mathcal{L}_{T_1 T_2 T_3}^{Con}$ .

that in our experiments, we do not observe significant changes in the overall performance for the weight hyper-parameter tuning, where  $\lambda = 0.3$  gives slightly better performance.

## 4. Experiments

### 4.1. Data

The dataset used in this study was collected from West China Hospital, Chengdu, China. A total of 4493 US images from 4493 patients (only one image was selected for a patient) are included in the dataset, in which 2576 images contain benign nodules and 1917 images contain malignant nodules. All images were acquired

using GE Logiq E9 ultrasound machine. For classification ground-truth annotations, all malignant cases have their corresponding FNA results and thus the classification labels are directly coming from the results. The classification labels of benign cases were verified by senior radiologists during clinical diagnosis. For segmentation ground-truth annotations, we follow the commonly-used annotation procedure of medical image segmentation datasets. Firstly, several junior radiologists delineated thyroid nodular region of each US image. Then, these delineation were checked and refined by three senior radiologists with more than 20 years clinical experiences. Finally, all refined segmentation annotations were examined again and agreed by all the three senior radiologists. The samples are excluded from the final dataset if there are variations among the three senior radiologists.

We random divide the dataset into training/validation/test subset with roughly 3:1:1 ratio. We follow stratified random sampling to ensure the distributions of benign and malignant in training/validation/test subsets are equal.

### 4.2. Evaluation metrics

We employ receiver operating characteristic (ROC), area under ROC curve (AUC), accuracy (ACC) and F1-score (F1) metrics to evaluate the performance of classification:

$$\begin{aligned}
 ACC &= \frac{TP + TN}{TP + TN + FP + FN} \\
 PRE &= \frac{TP}{TP + FP} \\
 REC &= \frac{TP}{TP + FN} \\
 F1 &= \frac{2 \times PRE \times REC}{PRE + REC},
 \end{aligned} \tag{10}$$

where TP, TN, FP, FN are the number of true positives, true negatives, false positives and false negatives, respectively.

For segmentation performance, we adopt Dice coefficient (denoted as Dice) and Intersection of Union (IoU):

$$\begin{aligned}
 Dice &= \frac{2 \times |S^{GT} \cap S^{pred}|}{|S^{GT}| + |S^{pred}|} \\
 IoU &= \frac{|S^{GT} \cap S^{pred}|}{|S^{GT} \cup S^{pred}|},
 \end{aligned} \tag{11}$$

where  $S^{GT}$  and  $S^{pred}$  are the segmentation ground truth and prediction, respectively. Considering we have two classes for nodule segmentation (benign and malignant), we further use mean Dice

**Table 1**  
Quantitative performance comparisons for thyroid nodule segmentation and classification.

Model	Segmentation						Classification		
	Dice			IoU			ACC	F1	AUC
	Mean	Benign	Malignant	Mean	Benign	Malignant			
ClsNet	–	–	–	–	–	–	0.8606	0.8318	0.9277
U-Net(Seg2) ⊕ ClsNet	0.7692	0.7391	0.7992	0.7299	0.6921	0.7677	0.8606	0.8318	0.9277
U-Net(Seg3) (Ronneberger et al., 2015)	0.7596	0.7286	0.7906	0.7180	0.6800	0.7559	–	–	–
U-Net+ (Zhou et al., 2019)	0.7582	0.7241	0.7922	0.7162	0.6756	0.7568	–	–	–
DeepLabv3+ (Chen et al., 2018a)	0.7484	0.7198	0.7770	0.7095	0.6749	0.7440	–	–	–
nnUNet (Isensee et al., 2021)	0.7605	0.7261	0.7949	0.7184	0.6761	0.7606	–	–	–
PSPNet (Zhao et al., 2017)	0.7346	0.6954	0.7737	0.6851	0.6374	0.7327	–	–	–
SegNet (Badrinarayanan et al., 2017)	0.6990	0.6579	0.7401	0.6524	0.6030	0.7017	–	–	–
Mask R-CNN (He et al., 2017)	0.7426	0.7111	0.7740	0.7009	0.6632	0.7386	–	–	–
Bi et al. (2017)	0.7373	0.6901	0.7844	0.6950	0.6406	0.7494	–	–	–
Kang et al. (2019)	0.7738	0.7435	0.8040	0.7320	0.6947	0.7693	–	–	–
MSL	0.7767	0.7448	0.8086	0.7364	0.6979	0.7749	–	–	–
MSL + Intra	0.7847	0.7520	0.8173	0.7442	0.7049	0.7835	–	–	–
Wang et al. (2018)	0.7708	0.7372	0.8044	0.7294	0.6897	0.7690	0.8744	0.8516	0.9398
Zhou et al. (2021)	0.7782	0.7515	0.8049	0.7385	0.7094	0.7675	0.8815	0.86087	0.9502
Chen et al. (2018b)	0.7561	0.7269	0.7852	0.7141	0.6784	0.7497	0.8718	0.8405	0.9423
Singh et al. (2019)	0.7548	0.7196	0.7900	0.7135	0.6715	0.7554	0.8774	0.8449	0.9426
MTL	0.7712	0.7386	0.8037	0.7295	0.6905	0.7685	0.8796	0.8568	0.9470
MTL + Intra	0.7928	0.7640	0.8215	0.7519	0.7173	0.7864	0.8952	0.8766	0.9562
MS-MTL	0.7813	0.7527	0.8099	0.7422	0.7082	0.7762	0.8785	0.8568	0.9523
<b>MS-MTL + Intra and Inter</b>	<b>0.8084</b>	<b>0.7819</b>	<b>0.8349</b>	<b>0.7675</b>	<b>0.7346</b>	<b>0.8003</b>	<b>0.9075</b>	<b>0.8915</b>	<b>0.9608</b>

**Table 2**

Different models in our experiments. CLS: classification, SEG: segmentation, NA: not applicable.

Model	Task	Task consistency
ClsNet	CLS	NA
U-Net(Seg2) ⊕ ClsNet	CLS + SEG	NA
U-Net(Seg3)	CLS+SEG	NA
MSL	SEG	Intra ( $\mathcal{L}_{T_2 T_3}^{Con}$ )
MTL	CLS + SEG	Inter ( $\mathcal{L}_{T_1 T_3}^{Con}$ )
MS-MTL	CLS + SEG	Intra and Inter ( $\mathcal{L}_{Consis}$ )

and mean IoU which are the average of corresponding metrics on the two classes.

Furthermore, in order to give a deep analysis of intra- and inter-task inconsistency of models with multiple homogeneous and heterogeneous tasks, we define three measurements to quantify the inconsistency at the task level:  $Inconsis_{cls}$  for the classification task,  $Inconsis_{seg2}$  for the 2-class segmentation task and  $Inconsis_{seg3}$  for the 3-class segmentation task. The corresponding calculations are defined as follows:

$$\begin{aligned}
 Inconsis_{cls} &= \frac{\text{Num}_{\text{ClsPred}}^{\text{Seg-as-Cls}}}{\text{Total number of test images}} \\
 Inconsis_{seg2} &= 1 - \text{IoU}(\text{Seg}_2^{\text{pred}}, \text{Seg}_3\text{-as-Seg}_2) \\
 Inconsis_{seg3} &= 1 - \text{IoU}(\text{Seg}_3^{\text{pred}}, \text{Seg}_2\text{-as-Seg}_3),
 \end{aligned} \quad (12)$$

where  $\text{Num}_{\text{ClsPred}}^{\text{Seg-as-Cls}}$  is the number of images whose Seg – as – Cls and classification predictions are not the same in the test set.

#### 4.3. Implementation details

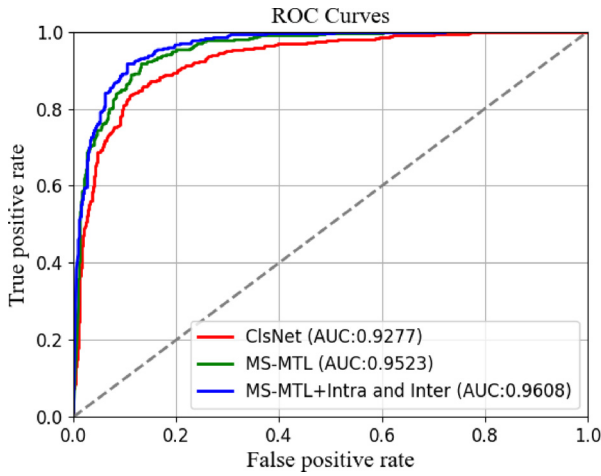
On the basis of MSL and MTL, we design a MS-MTL network as depicted in Fig. 2. Based on this network, we propose two different types of task consistency: intra- and inter-task consistency. In order to evaluate the effectiveness of each consistency component, as well as the MSL and MTL for thyroid nodule segmentation and classification, we construct several different models in our experiments. These models are listed in Table 2. Specifically,

we construct three baseline models: ClsNet for a single classification task, U-Net(Seg3) for a single 3-class segmentation task, and U-Net(Seg2) ⊕ ClsNet which is the combination of two separated models that are trained separately. Besides these baseline models, we further construct three models based on MSL and/or MTL. The MSL model is a multi-stage learning model which cascading two U-Nets, where the first U-Net in the first stage performs 2-class segmentation while the second U-Net in the second stage predicts 3-class segmentation. The MTL model is a multi-task learning model performing 3-class segmentation and classification in parallel as presented in Fig. 1(a). And finally the MS-MTL model is our designed network which is the combination of MSL and MTL as demonstrated in Fig. 2. For the convenience of notations, we use ‘+Intra’ and/or ‘+Inter’ after the models to denote the presence of intra-task and/or inter-task consistent learning.

Based on these designed models, three groups of comparative experiments are used to validate the effectiveness of each component in the task consistent learning. Firstly, since the two segmentation tasks in the MSL model are homogeneous pixel-wise tasks, the MSL model can be used to evaluate the intra-task consistency:  $\mathcal{L}_{T_2 T_3}^{Con}$ . Secondly, because the segmentation and classification in the MTL model are heterogeneous, it can be utilized to evaluate the inter-task consistency:  $\mathcal{L}_{T_1 T_3}^{Con}$ . Finally, the MS-MTL model performs two homogeneous segmentation tasks and one heterogeneous classification task; as a consequence, the MS-MTL model is used to evaluate the combination of intra- and inter-task consistency.

In order to make a fair comparison, we adopt ResNet101 (He et al., 2016) as the backbone for all of the models. Specifically, ResNet101 is used as the ClsNet model for classification. For all segmentation models, we employ U-Net with ResNet101 as the encoder. We implement all models with PyTorch (Paszke et al., 2019). The ADAM optimizer (Kingma and Ba, 2014) with a decaying learning rate initialized at  $3e-4$  is used to train all the models. Extensive data augmentation including flipping, affine transformation, rotation, Gaussian noise, random contrast and brightness are carried out during networks training. We trained all models on several NVIDIA V100 GPUs and the batch size is set to 16. The input image size for all the models is  $512 \times 512$ .





**Fig. 4.** ROC curves of different models including ClsNet, MS-MTL, and MS-MTL + Intra and Inter consistent learning.

**Table 3**  
Quantitative results of inconsistency in different models (%).

Model	$Inconsis_{seg2}$	$Inconsis_{seg3}$		$Inconsis_{cls}$
		Benign	Malignant	
MSL	3.43	–	–	–
MSL + Intra	1.86	–	–	–
MTL	–	–	–	14.49
MTL + Intra	–	–	–	6.80
MS-MTL	3.38	12.19	9.65	14.52
<b>MS-MTL + Intra &amp; Inter</b>	<b>1.23</b>	<b>4.92</b>	<b>4.05</b>	<b>3.68</b>

## 5. Results

### 5.1. Quantitative segmentation and classification results

Table 1 gives the quantitative performance comparison of different models on thyroid nodule segmentation and classification. Besides the models listed in Table 2, we also make a comparison with six commonly used medical segmentation networks: U-Net++ (Zhou et al., 2019), DeepLabv3+ (Chen et al., 2018a), nnUNet (Isensee et al., 2021), PSPNet (Zhao et al., 2017), SegNet (Badrinarayanan et al., 2017) and Mask R-CNN (He et al., 2017), all of them perform 3-class segmentation. Furthermore, two MSL solutions (Bi et al., 2017; Kang et al., 2019) and four MTL networks (Wang et al., 2018; Zhou et al., 2021; Chen et al., 2018b; Singh et al., 2019) are included for comparison. For the network proposed by Wang et al. (2018), we use the segmentation and classification part. For the model in Zhou et al. (2021), we convert all 3D operations to 2D operations and disable the iterative feature refinement since this strategy is a standalone training component that can be used with any segmentation networks. For fair comparison, we also use the ResNet-101 (He et al., 2016) as the backbone for all of the models. We will explain the table in details later in the subsections.

Fig. 4 displays the classification ROC curves of three different models: ClsNet, MS-MTL, and MS-MTL + Intra and Inter model. Compared with the single classification model ClsNet, the MS-MTL model improves the AUC from 0.9277 to 0.9523, and the MS-MTL model with the proposed intra- and inter-task consistency further increases AUC to 0.9608. With the inconsistency measurements defined in Section 4.2, we also compute the quantitative results of inconsistency on the three models (MSL, MTL and MS-MTL) with/without different components of task consistency. The results are shown in Table 3.

In general, from the performance results in Table 1 and the ROC curves in Fig. 4, the MS-MTL model with our proposed intra- and inter-task consistent learning (i.e., MS-MTL + Intra and Inter) achieves the best performance for both segmentation and classification. In addition, from the performance results of our designed three comparative experiments related with task consistency, the models with task consistent learning (i.e., MSL + Intra, MTL + Intra and MS-MTL + Intra and Inter) constantly perform much better compared with their corresponding baseline models without task consistency. In the following subsections, we respectively elaborate our results for the three groups of comparative experiments.

#### 5.1.1. Effectiveness of intra-task consistency in multi-stage learning

From the comparative experimental results of MSL and MSL + Intra models in Table 1, we can observe the effectiveness of our proposed intra-task consistency ( $\mathcal{L}_{T_2T_3}^{Con}$ ) for 3-class segmentation. Specifically, compared with the baseline MSL model for each segmentation metric, the MSL + Intra model improves mean Dice by 0.8% and mean IoU by 0.78%.

Moreover, from the 2-class segmentation inconsistency results of the MSL model and MSL + Intra model in Table 3, we can clearly see that by adding intra-task consistency in the MSL model, the  $Inconsis_{seg2}$  drops from 3.43% to 1.86%, which means that our proposed intra-task consistent learning can effectively eliminate the inconsistency between 2-class segmentation and 3-class segmentation.

Combining the above results and analysis, it indicates that our proposed intra-task consistency enforces the predictions of two homogeneous segmentation tasks to be consistent with each other and thus improves the segmentation performance.

#### 5.1.2. Effectiveness of inter-task consistency in multi-task learning

Similarly, comparing the performance results of MTL and MTL + Intra models shown in Table 1, we can see the effectiveness of our proposed inter-task consistency for both segmentation and classification. Specifically, compared with the baseline MTL model, our MTL + Intra model with one of the proposed inter-task consistency ( $\mathcal{L}_{T_1T_3}^{Con}$ ) increases mean Dice by 2.16% and mean IoU by 2.24% for the segmentation task. For the classification task, it improves F1 by 1.98% and AUC by 0.92%.

Additionally, from the inconsistency results of MTL and MTL + Intra model for classification shown in Table 3, we can also see that the  $Inconsis_{cls}$  is significantly decreased from 14.49% in the baseline MTL model to 6.80% in our proposed MTL + Intra model, which illustrates that the inter-task consistent learning can effectively reduce the inconsistency between 3-class segmentation and classification.

In summary, the aforementioned results and analysis demonstrate that our proposed inter-task consistent learning with  $\mathcal{L}_{T_1T_3}^{Con}$  can successfully decrease the inconsistent predictions between multi-class segmentation and classification, therefore improve the performance of both tasks.

#### 5.1.3. Effectiveness of intra- and inter-task consistency in multi-stage multi-task learning

Finally, the performance comparison of the MS-MTL model with/without intra- and inter-task consistent learning (Table 2) also shows that the combination of intra- and inter-task consistency can lead to the following improvements: 2.71% and 2.53% increase with respect to the mean Dice of benign and malignant class, respectively; 3.47% increase with respect to F1; and 0.85% increase with respect to AUC. These improvements in both segmentation and classification metrics again demonstrate the effectiveness of our proposed task consistency with the combination of intra- and inter-task consistency ( $\mathcal{L}_{Consis}$ ).

On the other hand, from the inconsistency results of MS-MTL and MS-MTL + Intra and Inter models in Table 3, we can observe the decline of all the three proposed inconsistency measurements, i.e., the  $Inconsis_{seg2}$  declines from 3.38% to 1.23%; the  $Inconsis_{seg3}$  for benign and malignant classes reduce from 12.19% and 9.65% to 4.92% and 4.05%, respectively; the  $Inconsis_{cls}$  decreases from 14.52% to 3.68%. These decrements show that the intra- and inter-task consistency can effectively eliminate the inconsistency for both segmentation and classification. Moreover, in terms of the inconsistency for 2-class segmentation, comparing with the MSL model with only intra-task consistency, the MS-MTL model with intra- and inter-task consistency achieves much lower  $Inconsis_{seg2}$ , which suggests the effectiveness of the combination of both intra- and inter-task consistent learning for reducing inconsistency between the two homogeneous segmentation tasks. Similarly, with respect to the inconsistency for the classification task, the combination of intra- and inter-task consistency also obtains much lower  $Inconsis_{cls}$  compared with the MTL model with only inter-task consistency, which again demonstrates the effectiveness of the combination of intra- and inter-task consistency for eliminating inconsistency between heterogeneous tasks, i.e., multi-class segmentation and classification.

Overall, the above analysis indicates that the combination of intra- and inter task consistency ( $\mathcal{L}_{Consis}$ ) can not only reduce the inconsistency between homogeneous segmentation tasks, but also decrease the inconsistency between heterogeneous segmentation and classification tasks, therefore increases the performance of each task.

## 5.2. Qualitative analysis

The visual performance comparison of different models for multi-class segmentation and classification is shown in Fig. 5. The first and second rows in Fig. 5 show that in all cases of the models that have right classification, our proposed MS-MTL + Intra and Inter model can produce better fine-grained contour segmentation. Notably, the inconsistency between multi-class segmentation and classification in the MS-MTL model is largely eliminated by our proposed task consistent learning. For example, the baseline MS-MTL model gives correct segmentation but wrong classification in the 3th and 4th rows in the figure, while the 5th and 6th rows have correct classification but wrong segmentation. To conclude, it can be observed that compared with other methods, our proposed MS-MTL + Intra and Inter model shows the best multi-class segmentation and classification results.

In order to give an intuition understanding of the effect after adding the task consistency, we compare the class activation map (CAM) (Zhou et al., 2016; Selvaraju et al., 2017) of classification predictions to the multi-class segmentation predictions based on the MS-MTL model with/without intra- and inter-task consistency. Specifically, we use the Grad-CAM (Selvaraju et al., 2017) to visualize the last convolution layer of classification branch in the MS-MTL model for the category predicted by the classification. In the first row of Fig. 6, although the MS-MTL model has correct segmentation (malignant), the CAM of classification is not concentrated on the nodule area and not consistent with the segmentation result; thus gives wrong classification result (benign). However, both of the two predictions are consistent and correct in the MS-MTL model with our task consistency. In the first row of Fig. 6, we can clearly observed that the CAM changes a lot when adopting task consistent learning. The main reason is that, for the models without consistency learning, the classification branch may predict wrong label (e.g., ground truth is malignant but predicted as benign like the first row of Fig. 6), resulting in the CAM not concentrated on the nodular region, but the segmentation branch can still locate the approximate nodular region and predict the category

**Table 4**

Ablation study results of task consistency. Seg: segmentation; Cls: classification.

Model	Task consistency			Seg Mean dice	Cls F1
	Intra-task		Inter-task		
	$\mathcal{L}_{T_2 T_3}^{Con}$	$\mathcal{L}_{T_1 T_3}^{Con}$	$\mathcal{L}_{T_1 T_2 T_3}^{Con}$		
MS-MTL				0.7813	0.8568
	✓			0.7887	0.8571
		✓		0.7944	0.8747
			✓	0.7961	0.8655
	✓	✓		0.8014	0.8745
	✓		✓	0.8028	0.8648
	✓	✓	✓	0.8045	0.8911
				<b>0.8084</b>	<b>0.8915</b>

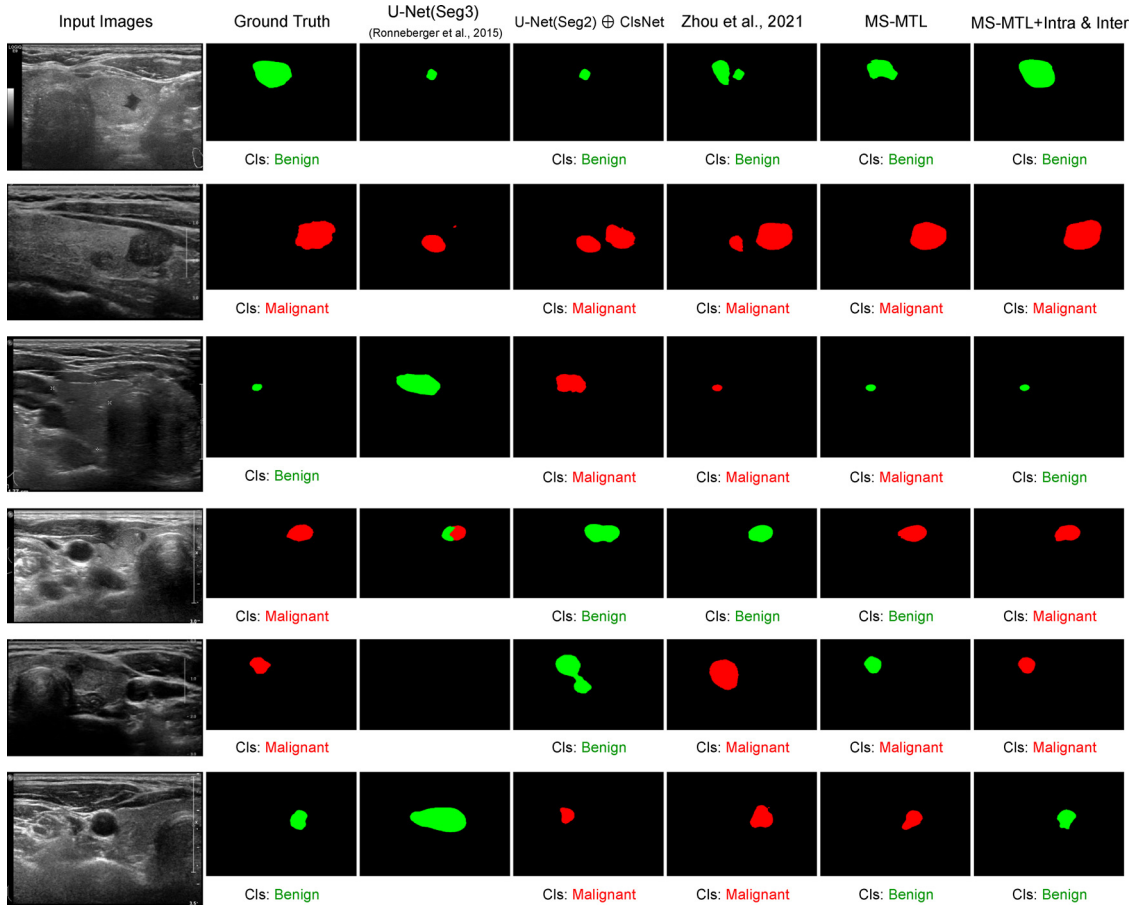
correctly (e.g., malignant nodule). In such cases, once the model is applied with consistency learning that enables the consistency between both of these two predictions, the segmentation branch can help to correct the classification prediction. This can be effectively reflected by the CAM of classification which is concentrated on the nodular region correctly. The second row of Fig. 6 shows that the MS-MTL model has correct classification where the CAM is roughly located on the nodule region, but gives inconsistent and wrong segmentation, which can be corrected by our proposed task consistency. To conclude, the two examples visually illustrate that our proposed task consistent learning enforces the MS-MTL model learn consistent predictions between classification and multi-class segmentation, which may coordinate with each other for improved results.

## 5.3. Ablation studies of task consistency

Although in Table 1, three comparative experiments related to each component of task consistency and their combination have already been explored, these experiments are done with different models, not with the same model. In order to further evaluate the performance contribution of each component of our proposed task consistency, i.e., one intra-task consistency ( $\mathcal{L}_{T_2 T_3}^{Con}$ ) and two kinds of inter-task consistency ( $\mathcal{L}_{T_1 T_3}^{Con}$  and  $\mathcal{L}_{T_1 T_2 T_3}^{Con}$ ), we conduct complete ablation study experiments based on our designed MS-MTL network as presented in Fig. 2 (i.e., the baseline is the MS-MTL model in Table 1).

### 5.3.1. Comparative experiments on each component in task consistency loss

Table 4 shows the ablation study results on each component of our proposed task consistency. First of all, by comparing ' $\mathcal{L}_{T_2 T_3}^{Con}$ ' in 'Intra-task' consistency and baseline (first row of Table 4), we can see that the intra-task consistency improves the performance for segmentation solely. Secondly, each component of 'Inter-task' consistency (i.e.,  $\mathcal{L}_{T_1 T_3}^{Con}$  and  $\mathcal{L}_{T_1 T_2 T_3}^{Con}$ ) both have better segmentation and classification performance compared with the baseline. Specifically, comparing these two inter-task components, the  $\mathcal{L}_{T_1 T_3}^{Con}$  achieves much lower segmentation performance but relatively higher classification performance, which reveals that the  $\mathcal{L}_{T_1 T_3}^{Con}$  focuses more on classification consistency, while the  $\mathcal{L}_{T_1 T_2 T_3}^{Con}$  emphasizes more on segmentation consistency. Thirdly, the combination of each component of the 'Inter-task' consistency with the 'Intra-task' consistency continuously improves the segmentation performance, but no further improvement is found on classification, which shows that the intra-task consistency has little effect on classification. Fourthly, the combination of two 'Inter-task' components performs much better for both segmentation and classification compared with each component alone, which proves the necessity of each



**Fig. 5.** Visual results of different models for multi-class segmentation and classification. From left to right: input images; ground truth for multi-class segmentation and classification; the results of U-Net (Ronneberger et al., 2015), U-Net(Seg2)⊕ClsNet, Zhou et al. (2021), MS-MTL and our MS-MTL + Intra and Inter consistent learning model. Benign nodules are shown in green and malignant nodules are shown in red.

Input Image	Ground Truth	MS-MTL		MS-MTL+Intra and Inter	
		Seg	CAM of Cls	Seg	CAM of Cls
			 Cls: Benign		 Cls: Malignant
			 Cls: Benign		 Cls: Benign

**Fig. 6.** Multi-class segmentation and class activation map (CAM) of classification in the MS-MTL model with/without intra- and inter-task consistency.

'Inter-task' component. Finally, the MS-MTL model with full components of intra- and inter-task consistency has the best segmentation and classification performance, which again justifies the effectiveness of each component and the combination of intra- and inter-task consistency.

### 5.3.2. Max-probability locations vs all-probability locations

Fig. 5 shows the performance comparison of using two different choices when calculating Seg – as – Cls for the inter-task consistency loss  $\mathcal{L}_{T_1 T_3}^{Con}$ , i.e., max-probability locations and all-probability locations. The MS-MTL is used as the base model. The performance comparison results clearly illustrate that the 'max-probability'

**Table 5**

Performance comparison of using max probability locations ('Max-probability') vs. all probability locations ('All-probability').

Model	Seg-as-Clis	Segmentation	Classification
		Mean dice	F1
MS-MTL	All-probability	0.7826	0.8576
	Max-probability	<b>0.7944</b>	<b>0.8747</b>

achieves much better performance than the 'all-probability' which indicates that using max-probability locations is more suitable for the computation of Seg – as – Cls.

**Table 6**

Ablation study results of the two different orders in the inter-task consistency  $\mathcal{L}_{T_1 T_3}^{Con}$ . The 'SC order' represents  $\mathcal{L}_{SCE}(Seg-as-Cls, Cls^{pred})$  and the 'CS order' means  $\mathcal{L}_{SCE}(Cls^{pred}, Seg-as-Cls)$  in Eq. (6). Seg: segmentation; Cls: classification.

Model	Inter-task consistency ( $\mathcal{L}_{T_1 T_3}^{Con}$ )		Seg	Cls
	SC order	CS order	Mean dice	F1
MS-MTL			0.7813	0.8568
	✓		0.7821	0.8708
		✓	0.7921	0.8563
	✓	✓	<b>0.7944</b>	<b>0.8747</b>

### 5.3.3. Orders in task consistency loss

In order to verify the effectiveness of two different orders which described in Eq. (6) in the inter-task consistency  $\mathcal{L}_{T_1 T_3}^{Con}$ . We conduct experiments based on the MS-MTL network and the results are shown in Table 6. The 'SC order' represents the  $\mathcal{L}_{SCE}(Seg-as-Cls, Cls^{pred})$  component and the 'CS order' means the  $\mathcal{L}_{SCE}(Cls^{pred}, Seg-as-Cls)$  component in Eq. (6). Specifically, the 'SC order' treats the segmentation results as ground truth and encourages the classification to be consistent with the segmentation while the 'CS order' treats the classification results as ground truth and enforces the segmentation to be consistent with the classification. From the performance results in Table 6, we can observe that the model with both of the two orders achieves the highest segmentation mean Dice and classification F1, which indicates the effectiveness of our choice in the inter-task consistency loss  $\mathcal{L}_{T_1 T_3}^{Con}$ .

## 6. Discussion

We propose a new paradigm of task consistency in MTL by combining intra- and inter-task consistency to eliminate the inconsistency among multiple homogeneous and heterogeneous tasks, therefore improving the performance for all related tasks. For the inconsistency between two homogeneous segmentation tasks, we propose intra-task consistency in which the Dice loss is utilized to enforce the two tasks learn consistent segmentation predictions in local fine-grained pixel level. On the other hand, for the inconsistency among heterogeneous segmentation and classification tasks, we propose inter-task consistency which contains two components, i.e., the consistency between classification and multi-class segmentation, and the consistency among all the three tasks (two segmentation tasks and one classification task). The consistency between classification and multi-class segmentation aims to enforce both of these two tasks learn consistent predictions in global coarse-grained image level, in other words, the class predictions of these two tasks should be the same. While the consistency among all the three tasks intends to enforce all the three tasks learn consistent predictions in both local fine-grained and global coarse-grained levels. Our experimental results verify the effectiveness of each component in the proposed task consistency. Moreover, since our proposed intra- and inter-task consistent learning is not at the model level, but at the task level, it can be applied to any MTL networks which may have multiple relevant intra- and inter-tasks. For the choice of network architectures, we choose U-Net as the main building block in this work due to its significant performance in medical image segmentation. However, any other segmentation networks can be used as the building block for the MTL network. Finally, it is also worth mentioning that although applied on thyroid nodule classification and segmentation tasks in ultrasound modality, we believe our designed solutions for measuring consistency are also suitable for other modalities (such as CT and MRI) and other tasks, depending on whether there exists evidence showing inconsistent predictions between different tasks and whether consistency constraints should be applied to reduce such conflicts.

Besides these contributions, our work mainly has one limitation for clinical usage. Since we construct our dataset following the same principle as the dataset which was used for thyroid nodule segmentation and classification in ultrasound images challenge of MICCAI 2020,<sup>1</sup> the US images in our dataset only contain benign or malignant nodules, i.e., the healthy thyroid cases (without nodules) and benign/malignant nodules co-exist cases are not included in our dataset, but these cases can be frequently found in clinical practice. However, for these cases, the task consistency proposed in this paper can still be used if extended with some modifications. For example, the intra-task consistency can be directly applied, while for the inter-task consistency, we can change the 2-class classification (benign and malignant) to a 4-class classification (normal, benign, malignant and co-exist) and make some extensions on the inter-task consistency. We will address this limitation in our future work.

## 7. Conclusion

In this paper, we apply multi-task learning for the joint thyroid nodule segmentation and classification in ultrasound images as a step towards thyroid nodule diagnosis in the computer-aided diagnosis systems. To overcome the inconsistency problems among multiple different but relevant tasks, we propose intra- and inter-task consistent learning to enforce the model learn consistent predictions among multiple homogeneous and heterogeneous tasks. Specifically, the intra-task part aims to solve the inconsistency among homogeneous tasks (pixel-wise segmentation tasks) while the inter-task part aims to tackle the inconsistency among heterogeneous tasks (pixel-wise segmentation task and categorical classification task). To further leverage the effectiveness of the proposed task consistent learning, we design a multi-stage multi-task learning (MS-MTL) network in which the first stage in the network performs binary segmentation and classification simultaneously and the second stage in the network learns multi-class segmentation.

We evaluate our proposed intra- and inter-task consistent learning and the designed MS-MTL network on a large thyroid ultrasound image dataset collected in West China hospital. The experimental results demonstrate that, on the basis of the MS-MTL network, the intra- and inter-task consistency improves the performance of both multi-class segmentation and classification tasks compared with the baseline MS-MTL model without task consistency. Specifically, the intra-task consistent learning declines the inconsistency between two homogeneous segmentation tasks and thus improves the segmentation performance in local fine-grained pixel level. On the other hand, the inter-task consistent learning decreases the inconsistency between heterogeneous segmentation and classification tasks, and therefore improves the multi-class segmentation and classification performance in global coarse-grained image level. Finally, the combination of intra- and inter-task consistency enforces consistent predictions in both local fine-grained pixel level and global coarse-grained image level across multiple homogeneous and heterogeneous tasks.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China under Grant 2020YFB1711500, and

<sup>1</sup> <https://sites.google.com/view/asmus2020>.



the 135 project for disciplines of excellence, West China Hospital, Sichuan University (ZYYC21004).

## References

- Alrubaidi, W.M., Peng, B., Yang, Y., Chen, Q., 2016. An interactive segmentation algorithm for thyroid nodules in ultrasound images. In: *International Conference on Intelligent Computing*. Springer, pp. 107–115.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12), 2481–2495.
- Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., Feng, D., 2017. Dermoscopic image segmentation via multitask fully convolutional networks. *IEEE Trans. Biomed. Eng.* 64 (9), 2065–2074.
- Bortsova, G., Dubost, F., Hogeweg, L., Katramados, I., de Bruijne, M., 2019. Semi-supervised medical image segmentation via learning consistency under transformations. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 810–818.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA* 68 (6), 394–424.
- Chang, C.Y., Chen, S.J., Tsai, M.F., 2010. Application of support-vector-machine-based method for feature selection and classification of thyroid nodules in ultrasound images. *Pattern Recognit.* 43 (10), 3494–3506.
- Chen, D., Zhang, X., Mei, Y., Liao, F., Xu, H., Li, Z., Xiao, Q., Guo, W., Zhang, H., Yan, T., et al., 2021. Multi-task learning for segmentation of aortic dissections using a prior aortic anatomy simplification. *Med. Image Anal.* 69, 101931.
- Chen, J., You, H., Li, K., 2020. A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images. *Comput. Methods Prog. Biomed.* 185, 105329.
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818.
- Chen, S.J., Chang, C.Y., Chang, K.Y., Tzeng, J.E., Chen, Y.T., Lin, C.W., Hsu, W.C., Wei, C.K., 2010. Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines. *Ultrasound Med. Biol.* 36 (12), 2018–2026.
- Chen, S., Wang, Z., Shi, J., Liu, B., Yu, N., 2018. A multi-task framework with feature passing module for skin lesion classification and segmentation. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, pp. 1126–1129.
- Chang, C.Y., Huang, H.C., Chen, S.J., 2009. Thyroid nodule segmentation and component analysis in ultrasound images. In: *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference, Asia-Pacific Signal and Information Processing Association, 2009 Annual*, pp. 910–917.
- Chi, J., Walia, E., Babyn, P., Wang, J., Groot, G., Eramian, M., 2017. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *J. Digit. Imaging* 30 (4), 477–486.
- Ding, J., Huang, Z., Shi, M., Ning, C., 2019. Automatic thyroid ultrasound image segmentation based on U-shaped network. In: *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, pp. 1–5.
- Du, W., Sang, N., 2015. An effective method for ultrasound thyroid nodules segmentation. In: *2015 International Symposium on Bioelectronics and Bioinformatics (ISBB)*. IEEE, pp. 207–210.
- Dwivedi, D., Aytar, Y., Tompson, J., Sermanet, P., Zisserman, A., 2019. Temporal cycle-consistency learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1801–1810.
- Haugen, B.R., Alexander, E.K., Bible, K.C., Doherty, G.M., Mandel, S.J., Nikiforov, Y.E., Pacini, F., Randolph, G.W., Sawka, A.M., Schlumberger, M., et al., 2016. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* 26 (1), 1–133.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969.
- He, K., Lian, C., Zhang, B., Zhang, X., Cao, X., Nie, D., Gao, Y., Zhang, J., Shen, D., 2021. HF-UNet: learning hierarchically inter-task relevance in multi-task U-Net for accurate prostate segmentation in CT images. *IEEE Trans. Med. Imaging* 40 (8), 2118–2128.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- He, T., Hu, J., Song, Y., Guo, J., Yi, Z., 2020. Multi-task learning for the segmentation of organs at risk with label dependence. *Med. Image Anal.* 61, 101666.
- Iakovidis, D.K., Keramidas, E.G., Maroulis, D., 2008. Fuzzy local binary patterns for ultrasound texture characterization. In: *International Conference Image Analysis and Recognition*. Springer, pp. 750–759.
- Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18 (2), 203–211.
- Kang, Q., Lao, Q., Fevens, T., 2019. Nuclei segmentation in histopathological images using two-stage learning. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 703–711.
- Katsigiannis, S., Keramidas, E.G., Maroulis, D., 2010. A contourlet transform feature extraction scheme for ultrasound thyroid texture classification. *Int. J. Eng. Intell. Syst. Electr. Eng. Commun.* 18 (3), 171.
- Keramidas, E.G., Maroulis, D., Iakovidis, D.K., 2012. TND: a thyroid nodule detection system for analysis of ultrasound images and videos. *J. Med. Syst.* 36 (3), 1271–1281.
- Kingma, D. P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Koundal, D., Gupta, S., Singh, S., 2012. Computer-aided diagnosis of thyroid nodule: a review. *Int. J. Comput. Sci. Eng. Surv.* 3 (4), 67.
- Koundal, D., Gupta, S., Singh, S., 2016. Automated delineation of thyroid nodules in ultrasound images using spatial neutrosophic clustering and level set. *Appl. Soft Comput.* 40, 86–97.
- Legakis, I., Savelonas, M.A., Maroulis, D., Iakovidis, D.K., 2011. Computer-based nodule malignancy risk assessment in thyroid ultrasound images. *Int. J. Comput. Appl.* 33 (1), 29–35.
- Li, X., Yu, L., Chen, H., Fu, C.W., Xing, L., Heng, P.-A., 2020. Transformation-consistent self-ensembling model for semisupervised medical image segmentation. *IEEE Trans. Neural Netw. Learn. Syst.* 32 (2), 523–534.
- Liu, T., Xie, S., Zhang, Y., Yu, J., Niu, L., Sun, W., 2017. Feature selection and thyroid nodule classification using transfer learning. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, pp. 1096–1099.
- Lu, Y., Pirk, S., Dlabal, J., Brohan, A., Pasad, A., Chen, Z., Casser, V., Angelova, A., Gordon, A., 2021. Taskology: utilizing task relations at scale. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8700–8709.
- Luo, X., Chen, J., Song, T., Wang, G., 2021. Semi-supervised medical image segmentation through dual-task consistency. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 8801–8809.
- Luo, S., Kim, E.H., Dighe, M., Kim, Y., 2011. Thyroid nodule classification using ultrasound elastography via linear discriminant analysis. *Ultrasonics* 51 (4), 425–431.
- Ma, J., Wu, F., Jiang, T., Zhao, Q., Kong, D., 2017. Ultrasound image-based thyroid nodule automatic segmentation using convolutional neural networks. *Int. J. Comput. Assist. Radiol. Surg.* 12 (11), 1895–1910.
- Ma, J., Wu, F., Zhu, J., Xu, D., Kong, D., 2017. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics* 73, 221–230.
- Maroulis, D.E., Savelonas, M.A., Iakovidis, D.K., Karkanis, S.A., Dimitropoulos, N., 2007. Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. *IEEE Trans. Inf. Technol. Biomed.* 11 (5), 537–543.
- Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, pp. 565–571.
- Mittal, S., Tatarchenko, M., Brox, T., 2019. Semi-supervised semantic segmentation with high- and low-level consistency. *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (4), 1369–1379.
- Nugroho, H.A., Nugroho, A., Choridah, L., 2015. Thyroid nodule segmentation using active contour bilateral filtering on ultrasound images. In: *2015 International Conference on Quality in Research (QIR)*. IEEE, pp. 43–46.
- Ouahabi, A., Taleb-Ahmed, A., 2021. Deep learning for real-time semantic segmentation: application in ultrasound imaging. *Pattern Recognit. Lett.* 144, 27–34.
- Ouali, Y., Hudelot, C., Tami, M., 2020. Semi-supervised semantic segmentation with cross-consistency training. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12674–12684.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32, 8026–8037.
- Playout, C., Duval, R., Chérét, F., 2018. A multitask learning architecture for simultaneous segmentation of bright and red lesions in fundus images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 101–108.
- Qu, H., Riedlinger, G., Wu, P., Huang, Q., Yi, J., De, S., Metaxas, D., 2019. Joint segmentation and fine-grained classification of nuclei in histopathology images. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, pp. 900–904.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Sajjadi, M., Javanmardi, M., Tasdizen, T., 2016. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Adv. Neural Inf. Process. Syst.* 29, 1163–1171.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626.
- Seo, H., Yu, L., Ren, H., Li, X., Shen, L., Xing, L., 2021. Deep neural network with consistency regularization of multi-output channels for improved tumor detection and delineation. *IEEE Trans. Med. Imaging* 40 (12), 3369–3378.
- Singh, V. K., Rashwan, H. A., Abdel-Nasser, M., Sarker, M., Kamal, M., Akram, F., Pandey, N., Romani, S., Puig, D., 2019. An efficient solution for breast tumor segmentation and classification in ultrasound images using deep adversarial learning. *arXiv preprint arXiv:1907.00887*.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9.

- Takahama, S., Kurose, Y., Mukuta, Y., Abe, H., Fukayama, M., Yoshizawa, A., Kitagawa, M., Harada, T., 2019. Multi-stage pathological image classification using semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10702–10711.
- Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780*.
- Tessler, F.N., Middleton, W.D., Grant, E.G., Hoang, J.K., Berland, L.L., Teefey, S.A., Cronan, J.J., Beland, M.D., Desser, T.S., Frates, M.C., et al., 2017. ACR thyroid imaging, reporting and data system (TI-RADS): white paper of the ACR TI-RADS committee. *J. Am. Coll. Radiol.* 14 (5), 587–595.
- Wang, P., Patel, V.M., Hacıhaliloglu, I., 2018. Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided CNN. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 134–142.
- Xie, X., Shi, F., Niu, J., Tang, X., 2018. Breast ultrasound image classification and segmentation using convolutional neural networks. In: *Pacific Rim Conference on Multimedia*. Springer, pp. 200–211.
- Ying, X., Yu, Z., Yu, R., Li, X., Yu, M., Zhao, M., Liu, K., 2018. Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network. In: *International Conference on Neural Information Processing*. Springer, pp. 373–384.
- Zamir, A.R., Sax, A., Cheerla, N., Suri, R., Cao, Z., Malik, J., Guibas, L.J., 2020. Robust learning through cross-task consistency. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11197–11206.
- Zhang, Y., Yang, Q., 2021. A survey on multi-task learning. *IEEE Trans. Knowl. Data Eng.* Early Access, 1.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890.
- Zhao, J., Zheng, W., Zhang, L., Tian, H., 2013. Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. *Health Inf. Sci. Syst.* 1 (1), 1–12.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929.
- Zhou, J., 2016. Thyroid tumor ultrasound image segmentation based on improved graph cut. In: *2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE, pp. 130–133.
- Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.-T., Shen, D., 2021. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Med. Image Anal.* 70, 101918.
- Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans. Med. Imaging* 39 (6), 1856–1867.
- Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232.
- Zou, Y., Luo, Z., Huang, J.-B., 2018. DF-Net: unsupervised joint learning of depth and flow using cross-task consistency. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 36–53.