DualNorm-UNet: Incorporating Global and Local Statistics for Robust Medical Image Segmentation

Junfei Xiao[1], Lequan Yu[2], Lei Xing[2], Alan L. Yuille[1], and Yuyin Zhou[2]

[1]Johns Hopkins University [2]Stanford University

Abstract

*Batch Normalization (BN) is one of the key components for accelerating network training, and has been widely adopted in the medical image analysis field. However, BN only calculates the global statistics at the batch level, and applies the same affine transformation uniformly across all spatial coordinates, which would suppress the image contrast of different semantic structures. In this paper, we propose to incorporate the semantic class information into normalization layers, so that the activations corresponding to different regions (i.e., classes) can be modulated differently. We thus develop a novel **DualNorm-UNet**, to concurrently incorporate both global image-level statistics and local region-wise statistics for network normalization. Specifically, the local statistics are integrated by adaptively modulating the activations along different class regions via the learned semantic masks in the normalization layer. Compared with existing methods, our approach exploits semantic knowledge at normalization and yields more discriminative features for robust segmentation results. More importantly, our network demonstrates superior ability in capturing domain-invariant information from multiple domains (institutions) of medical data. Extensive experiments show that our proposed DualNorm-UNet consistently improves the performance on various segmentation tasks, even in the face of more complex and variable data distributions. Code is available at* https://github.com/lambert-x/DualNorm-Unet.

1. Introduction

Medical image segmentation is an essential prerequisite for developing healthcare systems, which can largely benefit clinical applications such as disease diagnosis, surgical planning, and prognosis evaluation. Convolutional Neural Networks (CNNs), especially U-Net [30], have become the dominant approach on various medical image segmentation tasks [37, 28, 41, 15]. As one of the fundamental components for effectively training CNNs, Batch Normalization (BN) [14] plays a vital role in accelerating and stabilizing the network training procedure, and has been widely adopted in the medical imaging domain.

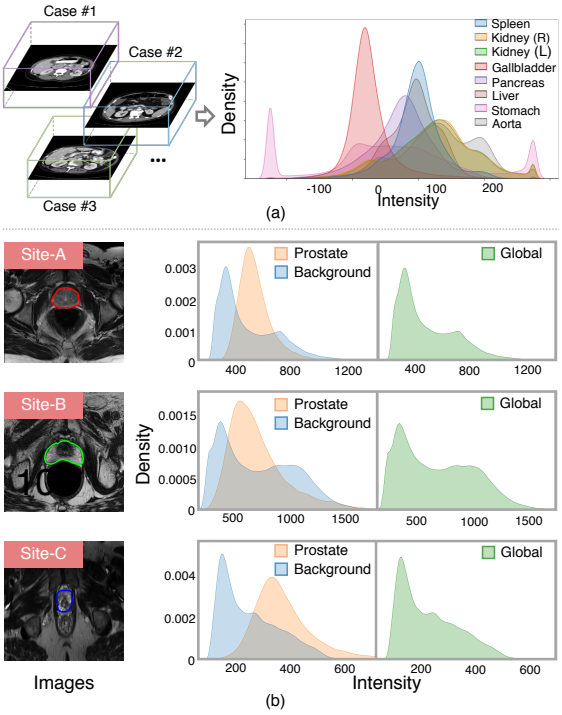

Figure 1. (a) region-wise intensity distribution of different subjects on abdominal CT images; (b) region-wise distribution vs. global distribution on different domains for prostate MRI images.

Meanwhile, various alternative normalization techniques have also been proposed for different usages. For instance, Group Normalization [35] was designed for small mini-batch settings, such as object detection and instance segmentation. Similarly, Layer Normalization [1] and Instance Normalization [33] were widely used for sequence models in natural language processing and image synthesis, respec-

tively. However, all these aforementioned normalization methods only estimate global statistics of the whole image while ignoring the local statistics of different semantic structures when conducting normalization operations. Considering that a typical medical image generally exhibits small intra-class variations and large inter-class variations (see Figure 1(a)) on the intensity distribution of different structures, we argue that such global statistics-based normalization strategies are suboptimal, especially for medical image segmentation. Moreover, in our preliminary experiments, we also found that these methods are not robust against heterogeneous data distributions of multi-domain medical images due to the bias of estimated global statistics [23]. As shown in Figure 1(b), the global batch-wise intensity distribution is quite similar to the background distribution, and thereby directly conducting BN on different domain images can lead to biased estimation for the foreground prostate region.

To alleviate these issues, we propose to jointly leverage local region-wise and global image-level statistics to conduct normalization, for enhancing medical image segmentation. On the one hand, the integration of local statistics can help derive more discriminative features of foreground structures, which can lead to a more refined segmentation outcome. On the other hand, we hypothesize that the integration of local statistics (*e.g.*, region-wise statistics) into normalization operations can better capture the domain-invariant information of different domain data, and thereby robustify the learned medical representations. As a proof test, we find that simply aligning the data distributions among different datasets through the region-wise statistics can already enhance the joint training performance (detailed in Section 3). This interesting observation further motivates us to explore the usage of local region-wise statistics, but in the normalization layers instead.

In this paper, we propose a new **DualNorm-UNet** for medical image segmentation, by introducing a novel *dual normalization* scheme to effectively integrate both local and global statistics into the training process. DualNorm-UNet aims to incorporate the semantic information in the normalization layer, so that the activations corresponding to different regions (*e.g.*, pancreas and stomach) can be modulated differently. Specifically, two parallel normalization streams are adopted for complementarily modulating the activation maps—one with the conventional BN for capturing the global statistics; while the other being a spatially-adaptive normalization follow the SPADE [29] strategy, which integrates local region-wise statistics with the guidance from the learned semantic masks. The proposed dual normalization scheme not only drives more discriminate feature extraction, but also can well compensate the statistical bias in the face of small distribution shifts or domain gaps.

We encapsulate the proposed dual normalization scheme in the residual block, referred to as *DualNorm Residual Block*, which forms the basic building block in the encoder of our DualNorm-UNet. As SPADE conditions on the input semantic masks, which are not available in the inference phase, we hereby introduce a simple yet effective multi-stage training strategy for addressing this issue. Concretely, we first train exclusively with the BN normalization for generating the semantic masks, and then feed them to SPADE for updating the spatially-adaptive modulating parameters. Compared with the existing normalization techniques, DualNorm-UNet achieves superior performances on various medical applications, including multi-organ segmentation from CT images and prostate segmentation from MRI images. In the face of multi-domain data, our method also demonstrates consistent improvement compared with state-of-the-arts. To conclude, our contributions are threefold:

- We propose a novel *dual normalization* scheme, for complementarily integrating local region-wise and global image-level statistics in the normalization operations during the training process for robust medical image segmentation.

- To our best knowledge, our method is the first to introduce SPADE in medical image segmentation. We further design a lightweight DualNorm Residual Block to be integrated into the U-Net encoder, which observes prominent performance gain with negligible additional parameters.

- Compared with existing normalization techniques, our DualNorm-UNet consistently achieves superior results, even with complex and variable data distributions.

2. Related Work

Medical image segmentation. Medical image segmentation is a crucial method for computer-aided diagnosis and has a wide range of targets including liver tumor [19, 34], brain tumor [10, 25], prostate [22, 37], abdominal organs [8, 40], *etc*. As Fully Convolutional Network [24] has shown remarkable performance on learning semantics with image representations, U-Net [30] builds upon this concept and proposes an encoder-decoder architecture with skip-connections which have been widely used as backbone for medical image segmentation. Based on this encoder-decoder architecture, V-Net [26] and 3D U-Net[4] use 3D convolution processing with 3D patches to fully extract context information in all 3 dimensions and some other methods[37, 28, 41, 15, 20, 3] are designed to introduce different mechanisms with achieving state-of-the-art performances on specific tasks.

Network normalization. Normalization is one of the keys to the success of deep networks. As the most commonly used normalization technique, Batch Normalization (BN) [14] enables training with larger learning rates and greatly mitigates general gradient issues. Besides, alternative normalization methods have also been designed for specific scenarios: Layer Normalization (LN) [1] for recurrent neural networks, Instance Normalization (IN) [33] for style transfer, Group Normalization (GN) [35] for small-batch training, *etc*. Beyond the natural image domain, these normalization methods have also been successfully applied to medical applications. For example, Kao *et al*. [16] apply GN for brain tumor segmentation; Isensee *et al*. [15] apply IN in a self-adaptive framework for various segmentation tasks; Chen *et al*. [3] apply LN in Transformers for abdominal multi-organ segmentation. A detailed comparison among different normalization for biomedical semantic segmentation and cross-modality synthesis have been summarized in [38] and [12]. To offer stronger affine transformation, later methods [6, 13] utilize external data to denormalize the features. As semantic information could be "washed away" with previous methods, SPADE [29] directly uses semantic masks for guiding the normalization. In addition, for domain adaptation and multi-domain learning, other methods propose to modify BN by modulating [21] or calculating domain-specific [2, 23] statistics. Different from these existing methodologies, this paper proposes a novel dual normalization scheme, for incorporating different types of statistics.

3. A Closer Look at Region-wise Statistics

To demonstrate the limitation of global statistics-based normalization, we sample three MRI images from different domains (institutions) and plot the global (batch-wise) vs. local (region-wise) intensity distributions as in Figure 1(b). Compared with the region-wise distribution, the batch-wise distribution is limited in capturing the image contrast between different classes. More critically, it can lead to biased estimation for certain classes. For instance, in Figure 1(b), we can observe that the batch-wise distribution is quite similar to the background distribution since the background occupies a relatively large portion of the entire image. Consequently, conducting normalization with only global statistics would lead to biased statistics estimation over the foreground prostate region.

We also design a simple experiment to investigate whether local statistics can better mitigate distribution shifts among different domains, compared to global statistics. Specifically, we simply align the input image distributions based on the region-/global- wise statistics obtained from different datasets. (More details of this algorithm can be found in the appendix.) Then we jointly train these heterogeneous datasets on the aligned input images. Our results

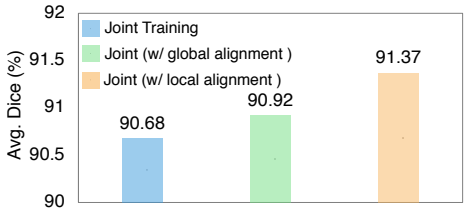

Figure 2. Training with heterogeneous prostate segmentation datasets under global/local alignment.

are summarized in Figure 2, which suggests that aligning the data distribution based on region-wise statistics can better mitigate the domain shifts among datasets and therefore achieves better results under the joint training setting. This interesting observation further motivates us to design better normalization strategies to further leverage the local statistics during the learning process, as detailed in Section 4.

4. Methodology

Given a set of input images $\mathbf{x} \in \mathbb{R}^D$ with the spatial resolution of D and the corresponding pixel-wise annotation masks $\mathbf{y} \in \mathbb{L}^D$, our goal is to predict the segmentation masks which can approach the groundtruth. To incorporate both the global and local semantic-aware statistics information into the normalization layer, we design a new *DualNorm residual block (DNRB)* module based on *dual normalization* scheme, which consists of Batch Normalization (BN) and Spatially-adaptive Normalization (SPADE). The proposed DNRB is further densely employed in the encoding path of the popular U-shaped architecture, *i.e*., U-Net, referred to as **DualNorm-UNet** throughout this paper. To facilitate the learning of such a dual normalization scheme, a multi-stage training paradigm is utilized. In the first stage, the network is trained with BN to generate the semantic masks, which are later fed to the second stage to facilitate the learning of SPADE. Figure 3 illustrates the overall pipeline of our approach. Below, we will first introduce our DNRB in Section 4.1. After that, the overall training and testing pipeline of DualNorm-UNet will be elaborated in Section 4.2.

4.1. DualNorm Residual Block (DNRB)

Different from standard residual blocks [11] where only Batch Normalization is used, we introduce two parallel normalization branches in our design while keeping all other components the same (see Figure 3). Let x with the spatial resolution of $H \times W$ denotes the input feature map, with a batch of N samples and C number of channels. For the n-th sample at the c-th channel, $x_{n,c,i,j}$ denotes the associated activation value at spatial location (i, j). Then a BN branch and a SPADE branch are used for fully leveraging the global and local statistics, respectively.

The BN branch. The BN [14] branch estimates the global

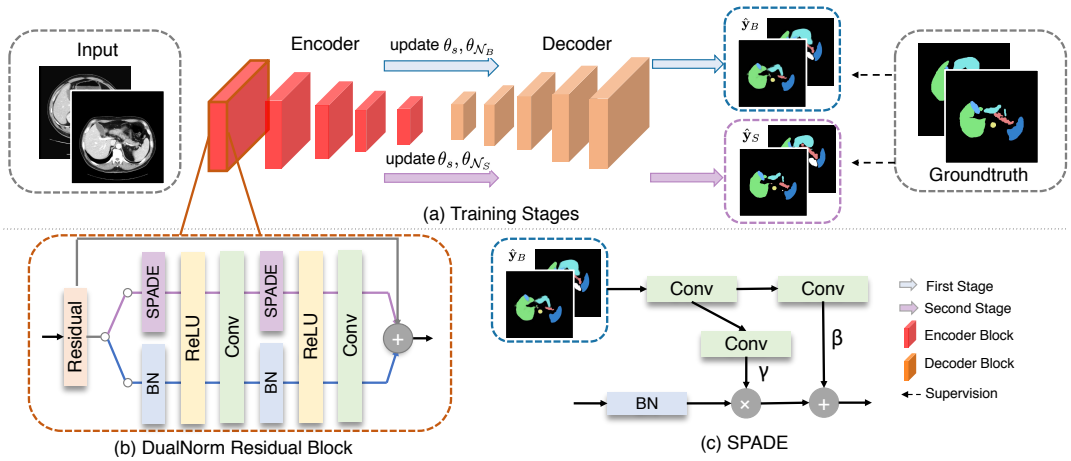

Figure 3. Overview of the framework. (a) architecture & training stages of DualNorm-UNet; (b) design of DualNorm Residual Block; (c) sketch of SPADE design.

statistics along the channel dimension and then applies affine transformation with learnable parameters γ_c^{BN} and β_c^{BN} at the c-th channel. Specifically, we first compute the mean and standard deviation μ_c and σ_c as follows:

$$\mu_c = \frac{1}{NHW}\sum_{n,i,j} x_{n,c,i,j}, \tag{1}$$

$$\sigma_c = \sqrt{\frac{1}{NHW}\sum_{n,i,j}\left((x_{n,c,i,j})^2 - (\mu_c)^2\right) + \epsilon}, \tag{2}$$

where ϵ denotes a small constant for avoiding invalid denominators. And the associated activation map can be then computed as:

$$\gamma_c^{\text{BN}} \cdot \frac{x_{n,c} - \mu_c}{\sigma_c} + \beta_c^{\text{BN}}. \tag{3}$$

The SPADE branch. We adopt SPADE [29] as an addition normalization to integrate local statistics. The key difference, compared with BN, is that SPADE provides a spatially-variant affine transformation which is learned from the corresponding semantic mask for modulating the activation map. Therefore, the modulation parameters $\gamma^{\text{SPADE}}(\mathbf{m})$ and $\beta^{\text{SPADE}}(\mathbf{m})$ are no longer C-dimensional vectors as aforementioned, but tensors with spatial resolution $H \times W$. Here $\gamma^{\text{SPADE}}(\cdot)$ and $\beta^{\text{SPADE}}(\cdot)$ are learnable functions which acts on the given semantic mask $\mathbf{m}$. The activations from the SPADE branch are computed as:

$$\gamma_{c,i,j}^{\text{SPADE}}(\mathbf{m})\frac{x_{n,c,i,j} - \mu_c}{\sigma_c} + \beta_{c,i,j}^{\text{SPADE}}(\mathbf{m}) \tag{4}$$

Following [29], we use a lightweight two-layer convolutional network to learn the modulation functions $\gamma^{\text{SPADE}}(\cdot)$ and $\beta^{\text{SPADE}}(\cdot)$, where the first layer is set to output features of $C/2$ number of channels.

The Choice of $\mathbf{m}$. In our problem, the semantic masks are not provided, unlike previous usages. Therefore, we use the pseudo-mask predicted by the BN branch as $\mathbf{m}$ for computing the modulation parameters $\gamma^{\text{SPADE}}(\mathbf{m})$ and $\beta^{\text{SPADE}}(\mathbf{m})$, and introduce a multi-stage training paradigm for learning the dual normalization accordingly in Section 4.2.

4.2. Learning with Dual Normalization

Our proposed DualNorm-UNet is denoted by $\mathcal{F}(\cdot;\theta)$ parameterized by $\theta = \{\theta_s, \theta_{\mathcal{N}_B}, \theta_{\mathcal{N}_S}\}$. $\theta_{\mathcal{N}_B}$ denote the learnable modulating parameters used in the BN branch and $\theta_{\mathcal{N}_S}$ is the learnable modulation function which requires the corresponding semantic mask as an input; the subscripts $\mathcal{N}_B$ and $\mathcal{N}_S$ here denote BN and SPADE normalization. And θ_s stands for all other network parameters.

Training with Batch Normalization. In the first stage, we exclusive train the model through the BN branch, *i.e.*, only θ_s and $\theta_{\mathcal{N}_B}$ are updated. The goal of the first stage is not only to leverage the global statistics for accelerating training but also to provide semantic information for learning SPADE parameters in the second stage. Specifically, the semantic mask generated from the first stage $\hat{\mathbf{y}}_B$ can be written as:

$$\hat{\mathbf{y}}_B = \mathcal{F}(\theta_s, \theta_{\mathcal{N}_B}; \mathbf{x}). \tag{5}$$

Training with Spatially-adaptive Normalization. In the second training stage, we train exclusively on the SPADE branch for exploiting the local statistics, *i.e.*, only update network parameters θ_s and the learnable function $\theta_{\mathcal{N}_S}(\cdot)$. Given the semantic mask $\hat{\mathbf{y}}_B$, the normalization parameters $\theta_{\mathcal{N}_S}(\hat{\mathbf{y}}_B)$ can be then computed via Equation 4, and the predicted mask in the second stage $\hat{\mathbf{y}}_S$ can be written as:

$$\hat{\mathbf{y}}_S = \mathcal{F}(\theta_s, \theta_{\mathcal{N}_S}(\hat{\mathbf{y}}_B); \mathbf{x}), \tag{6}$$

where $\hat{\mathbf{y}}_S$ denotes the softmax probability map output from the SPADE branch. With auxiliary semantic information integrated, this training step aims at enhancing discriminative

Algorithm 1 Training procedure of DualNorm-UNet

Input: Images and labels $\mathbf{x}, \mathbf{y}$;
Network parameters $\theta = \{\theta_s, \theta_{\mathcal{N}_B}, \theta_{\mathcal{N}_S}\}$;
Training iterations τ;
Output: Optimized parameters $\theta_s, \theta_{\mathcal{N}_B}, \theta_{\mathcal{N}_S}$;

t $\leftarrow 0$;
Initialize θ_s, $\theta_{\mathcal{N}_B}$ with the pretrained model and randomly initialize $\theta_{\mathcal{N}_S}$;
while $t < \tau$ **do**
  Compute the semantic mask $\hat{\mathbf{y}}_B$ via Equation (5);
  $\alpha \leftarrow 1$;
  Update $\theta_s, \theta_{\mathcal{N}_B} \leftarrow \min_{\theta_s, \theta_{\mathcal{N}_B}} \mathcal{L}_{total}$;
  Detach $\hat{\mathbf{y}}_B$ from gradient calculation;
  Compute the semantic mask $\hat{\mathbf{y}}_S$ via Equation (6);
  $\alpha \leftarrow 0$;
  Update $\theta_s, \theta_{\mathcal{N}_S}(\cdot) \leftarrow \min_{\theta_s, \theta_{\mathcal{N}_S}} \mathcal{L}_{total}$;
  $t \leftarrow t + 1$;
end while

features, which leads to more accurate and robust segmentation.

Overall training objective. Given a pair of probability prediction $\hat{\mathbf{y}}$ and the associated groundtruth $\mathbf{y} \in \mathbb{L}^D$, the Dice loss and the cross entropy loss are:

$$\mathcal{L}_{Dice}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{|\mathbb{L}|} \sum_{l} [1 - \frac{2 \sum_{i,j,l} y_{i,j,l} \cdot \hat{y}_{i,j,l}}{\sum_{i,j,l} (y^2_{i,j,l} + \hat{y}^2_{i,j,l})}], \quad (7)$$

$$\mathcal{L}_{CE}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{D \cdot |\mathbb{L}|} \sum_{i,j,l} y_{i,j,l} \cdot \log(\hat{y}_{i,j,l}), \quad (8)$$

where $\hat{y}_{i,j,l}$ is the output probability of the l-th class ($l \in \mathbb{L}$) of at spatial location i, j. In our loss function, we use a weighted sum of these two losses, which can be written as:

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \lambda \mathcal{L}_{Dice}(\mathbf{y}, \hat{\mathbf{y}}) + (1 - \lambda) \mathcal{L}_{CE}(\mathbf{y}, \hat{\mathbf{y}}), \quad (9)$$

where λ is the balance parameter. Therefore, our overall training objective over these two stages is:

$$\mathcal{L}_{total} = \alpha \mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}_B) + (1 - \alpha) \mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}_S), \quad (10)$$

where α is set as 1 in the first stage and 0 in the second stage, for updating $\{\theta_s, \theta_{\mathcal{N}_B}\}$ and $\{\theta_s, \theta_{\mathcal{N}_S}\}$ alternately. In the testing phase, the final prediction is obtained by forwarding twice with Equation (5) & (6) sequentially. The whole training procedure is summarized in Algorithm 1.

5. Experiments

5.1. Dataset

In order to demonstrate the effectiveness and the robustness of our approach, we conduct extensive experiments on prostate segmentation and multi-organ segmentation in both single-domain and multi-domain settings.

Dataset	#Cases	Field strength (T)	Resolution (in/through plane)	Manufacturer
ISBN-R	30	3	0.6-0.625/3.6-4	Siemens
ISBN-B	30	1.5	0.4/3	Philips
I2CVB	19	3	0.67-0.79/1.25	Siemens

Table 1. Details about the three prostate segmentation datasets.

Prostate segmentation datasets. Following [23], we use prostate T2-weighted MRI collected from three different domains, including 1) 30 samples from Radboud University Nijmegen Medical Centre; 2) 30 samples from Boston Medical Center; 3) 19 samples from Initiative for Collaborative Computer Vision Benchmarking (I2CVB) dataset [18]. 1) and 2) are available from NCI-ISBI 2013 challenge (ISBI 13) dataset [27], therefore are denoted as "ISBN-R" and "ISBN-B". And 3) will be denoted as "I2CVB". Details of their acquisition protocols are shown in Table 1.

Abdominal multi-organ segmentation datasets. We use abdominal CT images of two different domains, including 1) 30 training cases from the Beyond the Cranial Vault (BTCV) [17] dataset; 2) 41 cases of the Cancer Image Archive (TCIA) Pancreas-CT dataset [5, 31, 32], where the multi-class annotation can be acquired from [9]. For single-domain experimental settings, 8 organs (spleen, left kidney, right kidney, gallbladder, pancreas, liver, stomach and aorta) are evaluated following the settings in [7, 3]. For multi-domain experiments, right kidney and aorta are excluded and we evaluate the remaining 6 organs which are labeled on both datasets.

5.2. Experimental Setting

Data preprocessing. For the prostate datasets, we follow the preprocessing steps in [23] and implement center-cropping, out-of-mask slice clipping, resizing and Z-score normalization. For abdominal multi-organ segmentation, following [39], we firstly clip the image intensities with the soft tissue CT window range of [-125, 275] and then also apply out-of-mask slice clipping and Z-score normalization. All images are resized to 384×384. For data augmentation, horizontal flipping and rotation are applied during the training for prostate segmentation while only rotation is used for multi-organ segmentation.

Implementation details. The hyper-parameter λ in Equation (9) is set as 0.5 for abdominal multi-organ segmentation and 1.0 for prostate segmentation respectively. During the training process, we use the Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$ and the initial learning rate is set to 10^{-3} whose decay is scheduled by Plateau Scheduler[1](

[1] ReduceLROnPlateau Schduler - Pytorch Version

Method	Pretrained	Aug	Norm	AVG	Spleen	Kidney (R)	Kidney (L)	Gallbladder	Pancreas	Liver	Stomach	Aorta
Baseline	✗	✗	BN	75.56	92.00	81.92	84.45	50.76	48.89	95.07	68.25	83.16
Baseline	✗	✓	BN	77.83	91.86	84.36	86.99	53.92	51.18	95.11	75.24	83.97
Baseline	✓	✓	BN	78.98	93.43	84.68	87.98	54.54	55.14	95.35	76.19	84.49
Baseline	✓	✓	LN	69.80	80.19	74.58	82.57	52.98	34.80	92.43	58.85	81.97
Baseline	✓	✓	IN	78.96	91.93	83.70	87.83	54.29	55.08	95.23	78.72	84.93
Baseline	✓	✓	GN	78.50	89.84	85.64	86.26	55.14	55.70	94.71	75.98	84.71
Ours (block1)	✓	✓	DualNorm	79.33	93.07	85.66	88.26	52.79	54.34	**95.64**	77.38	**87.49**
Ours (block1-4)	✓	✓	DualNorm	**80.37**	**94.63**	**86.29**	**88.64**	**55.51**	**55.91**	**95.64**	**79.80**	86.52

Table 2. Comparison on the multi-organ segmentation dataset with single-domain setting (Dice Score in %).

Method	Pretrained	Aug	Norm	ISBN-R	ISBN-B	I2CVB	AVG
Baseline	✗	✗	BN	88.69	85.2	87.21	87.03
Baseline	✗	✓	BN	89.13	85.99	88.22	87.78
Baseline	✓	✓	BN	90.08	87.22	88.99	88.76
Baseline	✓	✓	LN	87.57	83.91	85.42	85.64
Baseline	✓	✓	IN	89.05	88.37	88.98	88.8
Baseline	✓	✓	GN	89.15	**88.61**	89.16	88.97
Ours (block 1)	✓	✓	DualNorm	91.36	88.55	89.31	**89.74**
Ours (block 1-4)	✓	✓	DualNorm	**91.41**	87.57	**89.79**	89.59

Table 3. Comparison on the prostate segmentation datasets with single-domain settings (Dice Score in %).

$factor = 0.5, patience = 5$) based on the average training loss of the last 50 iterations. The batch size is set as 4/6 for single-/multi- domain experiments respectively. The number of training iterations τ is set as 9K.

Network initialization. A pretraining stage of 18K iterations is applied for initializing θ_s and $\theta_{\mathcal{N}_B}$ of DualNorm-UNet. Meanwhile, to guarantee a fair comparison with others, we also apply the same initialization to all comparison methods.

Evaluation metric. Following the standard cross-validation evaluation [32, 40], we randomly split each dataset into 5 complementary folds, then apply 5-fold cross-validation. We use the Dice Coefficient and Average Symmetric Distance (ASD) to measure the segmentation performance.

5.3. Comparison with state-of-the-arts

In this section, we compare the proposed dual normalization scheme (denoted as "DualNorm") to various normalization methods, including BN [14], IN [33], GN [35], and LN [1]. In addition, for multi-domain settings, we also compare with state-of-the-art multi-domain learning approaches, including DSBN [2], MS-Net [23]. To ensure a fair comparison, we implement a Residual-UNet [23, 37] as the same backbone architecture for all methods.

Single-domain segmentation results. We summarize the single-domain segmentation results on three prostate segmentation datasets and one abdominal dataset as shown in Table 2 and 3. We can observe that even with strong data augmentation, the proposed dual normalization still yields a solid performance gain on all four datasets. For instance, on the prostate dataset "ISBN-R" and the multi-organ segmentation dataset, our DualNorm-UNet outperforms the BN counterpart by a large margin of 1.33% and 1.39% in average Dice. While different normalization methods (*e.g.*, GN, IN) may behave similarly, our dual normalization consistently achieves better results compared with all other methods, which validates the effectiveness of our approach.

We also compare two different configurations of DualNorm-UNet: 1) *DualNorm-UNet (block 1)* which only replaces the first encoder block as DNRB; 2) *DualNorm-UNet (block 1-4)* which replaces all of the first four encoder blocks as DNRBs. For prostate segmentation, we find that both variants show a solid improvement while *DualNorm-UNet (block 1)* with only few additional parameters performs similarly as *DualNorm-UNet (block 1-4)* (*i.e.*, 89.74% vs. 89.59%). On the contrary, for multi-organ segmentation, *DualNorm-UNet (block 1)* demonstrates inferior results than *DualNorm-UNet (block 1-4)* (*i.e.*, 79.33% vs. 80.37%). This suggests that more DNRBs can bring additional benefits for more complex tasks such as multi-organ segmentation. For the relatively simpler binary segmentation task, *DualNorm-UNet (block 1)* might be enough to learn a good model, therefore using more DNRBs does not lead to further performance gain. A more detailed study regarding where to add DNRBs will be illustrated in Section 5.4.

Multi-domain segmentation results. We also evaluate our method under the multi-domain setting as in [23]. To demonstrate the effectiveness of DualNorm-UNet, we compare the performance with the baseline and state-of-the-art multi-domain learning methods (*i.e.*, DSBN, MS-Net) on three prostate segmentation datasets and two multi-organ segmentation datasets. As shown in Table 4 and 5, under strong data augmentation (*e.g.*, rotation, flipping), DSBN and MS-Net do not yield improvements anymore, while our method still secures a reasonable improvement compared to the baseline. For instance, *DualNorm-UNet (block 1-4)* outperforms the baseline by 0.81% and 1.05% in average Dice on the BTCV and TCIA dataset respectively.

This indicates that, unlike previous methods which disentangle the normalization layers for different domains, our method can better distill domain-invariant information in the face of a more complex and variable data distribution via the proposed dual normalization. Meanwhile, it is also

Method	Forward	BTCV	TCIA	AVG	Spleen	Kidney (L)	Gallbladder	Liver	Stomach	Pancreas
Baseline	BN	82.64	87.33	84.98	95.22	93.54	69.63	96.01	85.52	69.97
DSBN [2]	BN	82.67	87.83	85.25	95.42	93.49	69.98	**96.16**	86.11	70.34
MS-Net [23]	BN	82.17	87.85	85.01	95.36	93.10	68.38	95.75	86.77	70.69
Ours (block1)	DualNorm	83.28	88.22	85.75	95.45	**93.63**	70.79	96.13	87.18	71.32
Ours (block1-4)	DualNorm	**83.45**	**88.38**	**85.92**	**95.55**	93.48	**71.53**	96.13	**87.20**	**71.60**

Table 4. Comparison on the multi-organ segmentation datasets under the multi-domain setting (Dice Score in %).

Method	Norm	ISBN-R	ISBN-B	I2CVB	AVG
Baseline	BN	91.50	90.46	90.34	90.76
DSBN [2]	BN	91.98	90.22	90.10	90.77
MS-Net [23]	BN	91.93	90.30	89.89	90.71
Ours (block1)	DualNorm	92.12	90.76	90.33	91.07
Ours (block1-4)	DualNorm	**92.47**	**91.17**	**90.79**	**91.48**

Table 5. Comparison on the prostate segmentation datasets under the multi-domain setting (Dice Score in %).

worth mentioning that our approach is complementary to previously domain-specific normalization methods. How to apply such our method and the domain-specific normalization jointly will be explored in the future study.

Unlike the single-domain setting, here we can see that *DualNorm-UNet (block 1-4)* outperforms *DualNorm-UNet (block 1)* for both prostate segmentation and multi-organ segmentation. We conjecture that this is due to that given a more complex data distribution, more DNRBs can bring additional benefits by imposing semantic guidance on the encoder more densely.

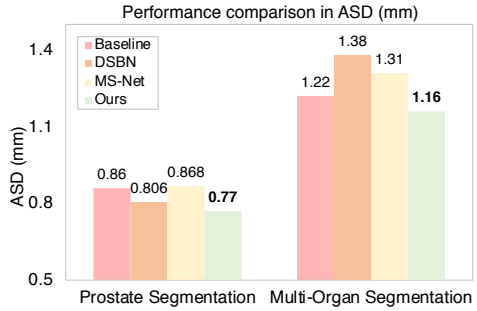

Figure 4. ASD (mm) comparison on both prostate segmentation and abdominal segmentation under multi-domain settings.

5.4. Ablation Study

Effectiveness of the Spatially-adaptive Normalization. Figure 5 visualizes the learned γ^{SPADE} and β^{SPADE} on different channels of the intermediate SPADE layers during the second forward. We can see that, with prior semantic information as guidance, SPADE can learn modulating parameters which are spatially-adaptive. Such spatial-wise modulation can be well complementary to the channel-wise modulation accomplished by BN, and derives more discriminative activations which can largely benefit medical image segmentation.

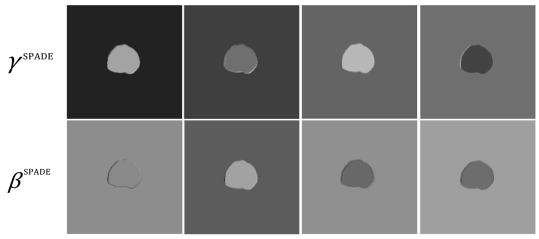

Figure 5. Visualization of the learned γ^{SPADE} (first row) and β^{SPADE} (second row) of a SPADE layer in a DualNorm block.

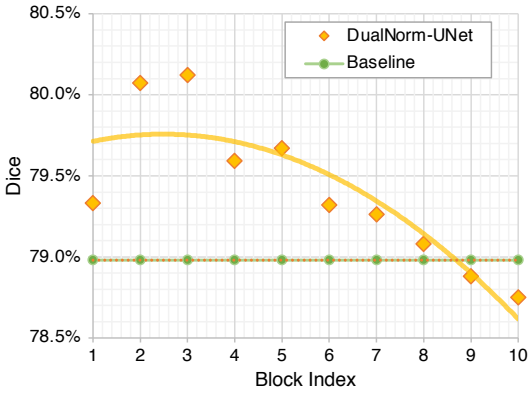

Figure 6. Single DualNorm block set in different position.

Where to add the DNRB? As aforementioned in Section 5.3, adding more DNRBs in the U-Net encoder can benefit both prostate segmentation and multi-organ segmentation, especially in multi-domain settings. This observation motivates us to further investigate where to add DNRB, as this can help us design better configurations of DualNorm-UNet which achieve higher performances without incurring much computation cost.

Our U-Net architecture consists of 5 encoder blocks and 5 decoder blocks. By varying the position to add the DNRB from block 1 to 10, we compare the average Dice score on the BTCV dataset. As shown in Figure 6, adding the DNRB to the encoder (block index 1-5) always yields better performance than adding to the decoder (block index 6-10).

In general, the performance decreases as the block index increases. We believe that this is due to that the earlier layers in the encoder are better at extracting low-level features which are less discriminative than the decoded fea-

tures. Therefore our dual normalization which incorporates semantic information naturally becomes a more favorable choice than BN for encoders.

Method	Sharing	#Params	Dice	ASD
Baseline	–	1.0×	90.08	0.71
Concatenated	✗	2.073×	90.36	0.72
Ours (block1)	✓	1.001×	91.36	0.63
Ours (block1-4)	✓	1.073×	**91.41**	**0.62**

Table 6. **Sharing** *vs*. **no-sharing**: we compare our method with *Concatenated*, where the network parameters in the two training stages are not shared.

	Forward	Prostate	Multi-Organ
Baseline	BN	90.76	84.98
Ours	1st (BN)	91.00	85.39
	2nd (SPADE)	**91.48**	**85.92**

Table 7. Performance with BN and SPADE in DualNorm block(s) in each forward (Dice Score in %).

Why sharing the rest of the network? To prove the necessity of network sharing except the normalization layers, we also implement our method by using different network parameters θ_s in the two training stages, similar to W-Net[36]. In this implementation, SPADE and BN are deployed in two independent sub-networks which are simply concatenated for training and testing (denoted as "Concatenated").

As shown in Table 6, our method performs much better than *Concatenated* in both average Dice and ASD with only about 50% of the parameters. Besides, even only comparing the results in the first stage where only BN is used during inference, as shown in Table 7, our approach still outperforms the baseline by 0.24% and 0.41% in average Dice. Then in the second stage where SPADE is used for inference, the performance can be further improved by 0.48% and 0.53%. This indicates that by sharing the rest of the network parameters, the two training stages can mutually benefit each other by leveraging both global-wise and local-wise normalization jointly.

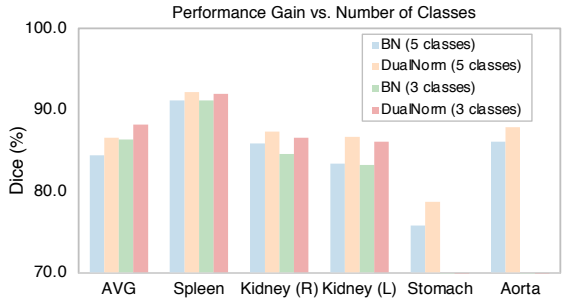

Figure 7. Performance gain under partial annotation.

The number of regions. Figure 7 compares our method to the baseline with fewer annotated classes (*i.e*., 3/5). We can see that by partitioning the images into different number of regions, *DualNorm* consistently achieves better results than BN for all tested organs. This suggests that our algorithm is not sensitive to the number of regions.

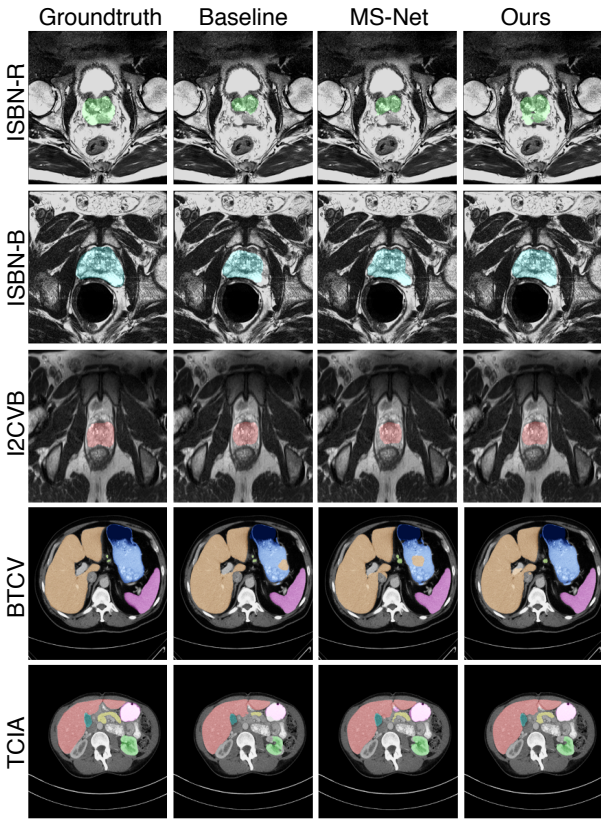

Figure 8. Qualitative Evaluation.

5.5. Qualitative Evaluation

Figure 8 qualitatively compares our method with the baseline and the other SOTA method under the multi-domain setting on prostate segmentation and abdominal multi-organ segmentation. Results in the first three rows clearly show that our method outperforms others as their results are cracked and incomplete with these unapparent prostate boundaries. And the results in the last two rows show our methods could better suppress inconsistent semantic information inside a close segmented area (*e.g*., reducing false positives inside the stomach) and predict hard organs like the pancreas more accurately by incorporating global and local statistics.

6. Conclusion

In this work, we propose a novel dual normalization scheme which complementarily integrates global and local statistics for robust medical image segmentation. To our best knowledge, our method is the first to introduce spatially-adaptive normalization in medical image segmentation, for capturing more discriminative and domain-invariant information. Compared with existing medical

image segmentation frameworks, our method consistently achieves superior results, even with complex and variable data distributions. In the future, we will study how to combine our method with the domain-specific normalization framework to further improve multi-domain learning, for both medical imaging and natural image domains.

Acknowledgement. We would like to thank Quande Liu for the discussion.

References

[1] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016. 1, 3, 6

[2] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7354–7362, 2019. 3, 6, 7, 11, 12

[3] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 2, 3, 5

[4] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016. 2

[5] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013. 5

[6] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629*, 2016. 3

[7] Shuhao Fu, Yongyi Lu, Yan Wang, Yuyin Zhou, Wei Shen, Elliot Fishman, and Alan Yuille. Domain adaptive relational reasoning for 3d multi-organ segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 656–666. Springer, 2020. 5

[8] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P Pereira, Matthew J Clarkson, and Dean C Barratt. Automatic multi-organ segmentation on abdominal ct with dense v-networks. *IEEE transactions on medical imaging*, 37(8):1822–1834, 2018. 2

[9] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P. Pereira, Matthew J. Clarkson, and Dean C. Barratt. Multi-organ Abdominal CT Reference Standard Segmentations, Feb. 2018. 5

[10] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[12] Shengye Hu, Jianpeng Yuan, and Shuqiang Wang. Cross-modality synthesis from mri to pet using adversarial u-net with different normalization. In *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, pages 1–5. IEEE, 2019. 3

[13] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017. 3

[14] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 1, 3, 6

[15] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. No new-net. In *International MICCAI Brainlesion Workshop*, pages 234–244. Springer, 2018. 1, 2, 3

[16] Po-Yu Kao, Thuyen Ngo, Angela Zhang, Jefferson W Chen, and BS Manjunath. Brain tumor segmentation and tractographic feature extraction from structural mr images for overall survival prediction. In *International MICCAI Brainlesion Workshop*, pages 128–141. Springer, 2018. 3

[17] B Landman, Z Xu, J Igelsias, M Styner, T Langerak, and A Klein. 2015 miccai multi-atlas labeling beyond the cranial vault workshop and challenge. 2015. 5

[18] Guillaume Lemaître, Robert Martí, Jordi Freixenet, Joan C Vilanova, Paul M Walker, and Fabrice Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: a review. *Computers in biology and medicine*, 60:8–31, 2015. 5

[19] Wen Li et al. Automatic segmentation of liver tumor in ct images with deep convolutional neural networks. *Journal of Computer and Communications*, 3(11):146, 2015. 2

[20] Xiaomeng Li, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE transactions on medical imaging*, 37(12):2663–2674, 2018. 2

[21] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016. 3

[22] Geert Litjens, Robert Toth, Wendy van de Ven, Caroline Hoeks, Sjoerd Kerkstra, Bram van Ginneken, Graham Vincent, Gwenael Guillard, Neil Birbeck, Jindang Zhang, et al. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. *Medical image analysis*, 18(2):359–373, 2014. 2

[23] Quande Liu, Qi Dou, Lequan Yu, and Pheng Ann Heng. Ms-net: Multi-site network for improving prostate segmentation with heterogeneous mri data. *IEEE Transactions on Medical Imaging*, 2020. 2, 3, 5, 6, 7, 11, 12

[24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 2

[25] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 2

[26] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016. 2

[27] Bloch N, Madabhushi A, Huisman H, Freymann J, Kirby J, Enquobahrie A Grauer M, Jaffe C, Clarke L, and Farahani K. Nci-isbi 2013 challenge: Automated segmentation of prostate structures., 2015. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.zF0vlOPv. 5

[28] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018. 1, 2

[29] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019. 2, 3, 4

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 1, 2

[31] Holger Roth, Amal Farag, Evrim B. Turkbey, Le Lu, Jiamin Liu, and Ronald M. Summers. Data from pancreas-ct, 2016. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.tNB1kqBU. 5

[32] Holger R Roth, Le Lu, Amal Farag, Hoo-Chang Shin, Jiamin Liu, Evrim B Turkbey, and Ronald M Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 556–564. Springer, 2015. 5, 6

[33] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 1, 3, 6

[34] R Vivanti, A Ephrat, L Joskowicz, O Karaaslan, N Lev-Cohain, and J Sosna. Automatic liver tumor segmentation in follow-up ct studies using convolutional neural networks. In *Proc. Patch-Based Methods in Medical Image Processing Workshop*, volume 2, 2015. 2

[35] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 1, 3, 6

[36] Xide Xia and Brian Kulis. W-net: A deep model for fully unsupervised image segmentation. *arXiv preprint arXiv:1711.08506*, 2017. 8

[37] Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng Ann Heng. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017. 1, 2, 6

[38] Xiao-Yun Zhou and Guang-Zhong Yang. Normalization in training u-net for 2-d biomedical semantic segmentation. *IEEE Robotics and Automation Letters*, 4(2):1792–1799, 2019. 3

[39] Yuyin Zhou, Zhe Li, Song Bai, Chong Wang, Xinlei Chen, Mei Han, Elliot Fishman, and Alan L Yuille. Prior-aware neural network for partially-supervised multi-organ segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10672–10681, 2019. 5

[40] Yuyin Zhou, Lingxi Xie, Wei Shen, Yan Wang, Elliot K Fishman, and Alan L Yuille. A fixed-point model for pancreas segmentation in abdominal ct scans. In *International conference on medical image computing and computer-assisted intervention*, pages 693–701. Springer, 2017. 2, 6

[41] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018. 1, 2

A. Aligning Input Distributions

Assume that we have N source domains $S_1, S_2, S_3, ..., S_N$, with $M_1, M_2, M_3, ..., M_N$ examples respectively, where the i-th domain source domain S_i consists of an image set $\{\mathbf{x}_{i,j} \in \mathbb{R}^{D_{i,j}}\}_{j=1,...,M_i}$ as well as their associated annotations. Our goal is to align the image distributions of these source domains with the target domain T based on the class-wise (region-wise) statistics. The algorithm can be illustrated as the following steps:

Step 1: Calculate class-wise statistics of each case

Firstly, we calculate the mean and standard deviation of each case in both the source domain and the target domain.

$$\mu_{i,j}^c = \frac{\sum_{k=1}^{|D_{i,j}^c|} \mathbf{x}_{i,j,k}^c}{|D_{i,j}^c|}, \tag{11}$$

$$\sigma_{i,j}^c = \sqrt{\frac{1}{|D_{i,j}^c|} \sum_{k=1}^{|D_{i,j}^c|} (\mathbf{x}_{i,j,k}^c - \mu_{i,j}^c)^2}, \tag{12}$$

where $\mathbf{x}_{i,j}^c$ denotes the pixels which belong to the c-th class (region) in image $\mathbf{x}_{i,j}$, with the number of pixels denoted as $|D_{i,j}^c|$. As a special case, $i = T$ indicates the target domain.

Step 2: Estimate aligned (new) class-wise statistics

Next, we calculate the mean of the statistics over all examples obtained in each domain as follows:

$$\bar{\mu}_i^c = \frac{\sum_{j=1}^{M_i} \mu_{i,j}^c}{M_i}, \tag{13}$$

$$\bar{\sigma}_i^c = \frac{\sum_{j=1}^{M_i} \sigma_{i,j}^c}{M_i}. \tag{14}$$

Based on the $\bar{\mu}_i^c$, we now estimate the new class-wise mean $\tilde{\mu}_{i,j}$ for each case of the source domain S_i as follows:

$$\tilde{\mu}_{i,j}^c = \frac{\mu_{i,j}^c - \bar{\mu}_i^c}{\sqrt{\frac{\sum_{j=1}^{M_i}(\mu_{i,j}^c - \bar{\mu}_i^c)^2}{M_i}}} \cdot \sqrt{\frac{\sum_{j=1}^{M_T}(\mu_{T,j}^c - \bar{\mu}_T^c)^2}{M_T}} + \bar{\mu}_T^c, \tag{15}$$

where M_T denotes the number of cases in the target domain T. Similarly, the new standard deviation $\tilde{\sigma}_{i,j}$ can be computed by:

$$\tilde{\sigma}_{i,j}^c = \frac{\sigma_{i,j}^c - \bar{\sigma}_i^c}{\sqrt{\frac{\sum_{j=1}^{M_i}(\sigma_{i,j}^c - \bar{\sigma}_i^c)^2}{M_i}}} \cdot \sqrt{\frac{\sum_{j=1}^{M_T}(\sigma_{T,j}^c - \bar{\sigma}_T^c)^2}{M_T}} + \bar{\sigma}_T^c. \tag{16}$$

Step 3: Align each case with the estimated statistics

Based on the computed new mean and standard deviation $\tilde{\mu}_{i,j}$, $\tilde{\sigma}_{i,j}$, the aligned image $\tilde{\mathbf{x}}_{i,j}$ can be computed as:

$$\tilde{\mathbf{x}}_{i,j}^c = \frac{\mathbf{x}_{i,j}^c - \mu_{i,j}^c}{\sigma_{i,j}^c} \cdot \tilde{\sigma}_{i,j}^c + \tilde{\mu}_{i,j}^c. \tag{17}$$

Method	Norm	ISBN-R	ISBN-B	I2CVB	AVG
Baseline	BN	0.64	0.71	1.22	0.86
DSBN [2]	BN	0.56	0.69	1.17	0.81
MS-Net [23]	BN	0.58	0.70	1.32	0.87
Ours (block1)	DualNorm	0.63	0.66	1.27	0.85
Ours (block1-4)	DualNorm	**0.54**	**0.64**	**1.13**	**0.77**

Table 8. ASD comparison on prostate segmentation datasets under the multi-domain setting (in mm). Compared with the baseline and other competitive methods, our proposed DualNorm-UNet achieves the lowest average ASD.

B. ASD Comparison

The detailed ASD comparison on both prostate segmentation and multi-organ segmentation can be found in Table 8 and 9. Our proposed DualNorm-UNet achieves the lowest average ASD on both tasks, even under the more challenging multi-domain setting.

C. Qualitative Evaluation

We also include more qualitative results on both prostate segmentation and multi-organ segmentation in Figure 9, where we show improved regions such as the gallbladder and the stomach compared with the baseline and other state-of-the-art multi-domain learning approaches [2, 23].

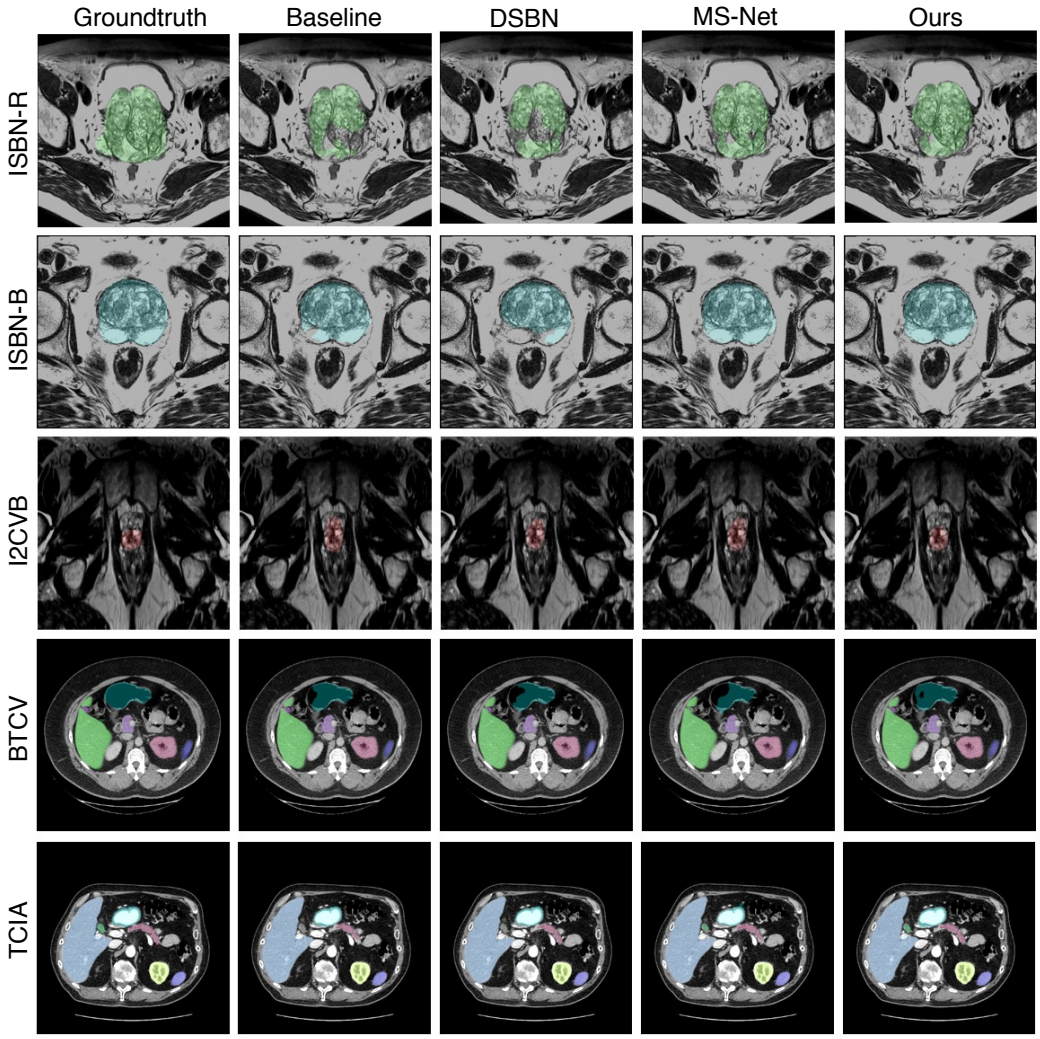

Figure 9. Qualitative Evaluation. For multi-organ segmentation, we can see improved regions such as the gallbladder and the stomach compared with the baseline and other state-of-the-art multi-domain learning approaches [2, 23].

Method	Forward	BTCV	TCIA	AVG	Spleen	Kidney (L)	Gallbladder	Liver	Stomach	Pancreas
Baseline	BN	1.28	1.17	1.22	0.59	0.59	2.36	0.77	1.93	1.10
DSBN [2]	BN	1.86	**0.90**	1.38	0.51	0.79	3.07	0.76	1.96	1.19
MS-Net [23]	BN	1.61	1.02	1.31	0.52	0.75	2.91	0.91	**1.58**	1.21
Ours (block1)	DualNorm	**1.22**	1.10	**1.16**	0.54	0.58	**2.22**	**0.74**	1.76	1.10
Ours (block1-4)	DualNorm	1.64	0.97	1.30	**0.51**	**0.55**	3.25	0.75	1.75	**1.01**

Table 9. ASD comparison on multi-organ segmentation datasets under the multi-domain setting (in mm). Comparedn with the baseline and other competitive methods, our proposed DualNorm-UNet achieves the lowest average ASD.