CSS-Net: Classification and Substitution for Segmentation of Rotator Cuff Tear

Kyungsu Lee[1], Hah Min Lew[2], Moon Hwan Lee[1], Jun-Young Kim[3(✉)], and Jae Youn Hwang[1(✉)]

[1] Daegu Gyeongbuk Institute of Science and Technology, Daegu 42988, South Korea
{ks_lee,moon2019,jyhwang}@dgist.ac.kr
[2] KLleon R and D Center, Seoul 04637, South Korea
hahmin.lew@klleon.io
[3] Daegu Catholic University School of Medicine, Daegu 42472, South Korea
dr.junyoung@gmail.com

Abstract. Magnetic resonance imaging (MRI) has been popularly used to diagnose orthopedic injuries because it offers high spatial resolution in a non-invasive manner. Since the rotator cuff tear (RCT) is a tear of the supraspinatus tendon (ST), a precise comprehension of both is required to diagnose the tear. However, previous deep learning studies have been insufficient in comprehending the correlations between the ST and RCT effectively and accurately. Therefore, in this paper, we propose a new method, *substitution learning*, wherein an MRI image is used to improve RCT diagnosis based on the knowledge transfer. The *substitution learning* mainly aims at segmenting RCT from MRI images by using the transferred knowledge while learning the correlations between RCT and ST. In substitution learning, the knowledge of correlations between RCT and ST is acquired by substituting the segmentation target (RCT) with the other target (ST), which has similar properties. To this end, we designed a novel deep learning model based on multi-task learning, which incorporates the newly developed substitution learning, with three parallel pipelines: (1) segmentation of RCT and ST regions, (2) classification of the existence of RCT, and (3) substitution of the ruptured ST regions, which are RCTs, with the recovered ST regions. We validated our developed model through experiments using 889 multi-categorical MRI images. The results exhibit that the proposed deep learning model outperforms other segmentation models to diagnose RCT with $6 \sim 8\%$ improved IoU values. Remarkably, the ablation study explicates that substitution learning ensured more valid knowledge transfer.

1 Introduction

In modern society, owing to the frequent incidence of rotator cuff tears (RCT) that occur in the supraspinatus tendon (ST) of people regardless of their age,

L. Wang et al. (Eds.): ACCV 2022, LNCS 13846, pp. 101–114, 2023.
https://doi.org/10.1007/978-3-031-26351-4_7

the demand for orthopedic diagnosis and surgery has increased recently [1]. RCT ruptures the shoulder joint, hindering movement of the shoulder [2,3]. However, to minimize resections, it is required to comprehend the precise location and size of RCTs and the mechanism behind them, prior to a surgical operation [4]. To this end, magnetic resonance imaging (MRI) has been established as an indispensable imaging tool owing to its non-invasive diagnostic capability to provide detailed anatomic structures. Using MRI, skilled surgeons have been able to localize RCTs and comprehensively analyze the tear. However, inter-clinician reliability and time-consuming manual segmentation have produced limitations in MRI-based diagnosis [5,6]. In contrast, advances in artificial intelligence have promoted the utilization of computer-assisted diagnosis (CAD) system in the medical imaging field [7–9]. Particularly, deep learning-based RCT diagnosis has been studied for the precise diagnosis of RCTs in terms of classification and segmentation. Kim *et al.* [10] detected the existence of RCT and classified the sizes, particularly a partial or full tear by adopting weighted combination layers. Shim *et al.* [11] exploited 3D CNN on volumetric MRI data to classify the existence of RCT and visualized the location of RCT using a gradient-weighted class activation mapping (Grad-CAM) [12].

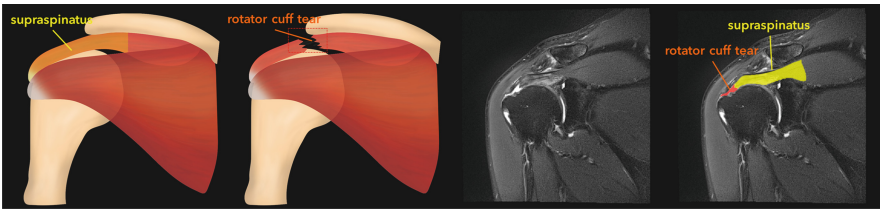

Fig. 1. Anatomical structure and MRI images of ST and RCT.

However, as illustrated in Fig. 1, the RCT region occupies a significantly smaller number of pixels in MRI images than the ST region, thus resulting in *class-imbalance problem*. Since the non-diseased regions correspond to most of the pixels, the trained network is biased toward the normal regions and converges to local minima [13,14]. In addition, since the RCT regions are sparse, the deep learning models could not learn enough knowledge related to RCT. To this end, researchers have used two major strategies; (1) a model-centric approach and (2) a data-centric approach. A novel loss function or network is proposed for the model-centric approach to resolving the biased state. The focal loss proposed by Lin *et al.* [15] was applied to resolve the class imbalance problem by assigning weights to the imbalanced class. Lee *et al.* [16] proposed a modified loss function to mitigate class imbalance in the diagnosis of RCTs from ultrasound images. However, previous studies and conventional algorithms for the class-imbalance problem have exhibited limited performance due to class imbalanced problem. Since these model-centric approaches could not

dramatically improve the accuracy with low-quality datasets, the data-centric approach should be accompanied by the model-centric approaches [17]. Recently, generative adversarial neural networks (GANs) [18] have been proposed as a useful tool for a data-centric approach. Several studies have demonstrated that data augmentation using GANs improves the accuracy of the diagnosis [19,20]. Particularly, data augmentation using GANs that mask lesions in synthetic images has been shown to be very useful even for medical applications [21–23]. However, because these synthetic images are not completely accurate, their use in the medical field remains debatable.

Therefore, to ensure the reliability of the generated medical images and improve diagnostic accuracy despite the class-imbalance problem, we propose a novel learning method of substitution learning for image translation as well as the corresponding network, denoted as the classification, substitution, and segmentation Network (CSS-Net). Initially, to ensure the generation of reliable medical images compared to GANs, substitution learning is newly developed using Discrete Fourier Transform (DFT) in the CSS-Net. Next, to improve the class-imbalance problem, wherein the knowledge related to RCT is limited due to sparse RCT information, we adopted the knowledge transfer-based method to CSS-Net to learn abundant knowledge of RCT from other tasks and other related classes. Since RCT is originally a part of ST and the RCT is meanwhile given from the tear of ST, there should be correlations between RCT and ST. At this moment, we were motivated that the knowledge about correlations between RCT and ST could be informative for other tasks, such as segmentation of RCT.

To this end, we designed the multi-task learning-based deep learning network, including segmentation and classification tasks. The simple transfer learning-based network could not still improve the segmentation accuracy drastically. Therefore, we were motivated to use image translation to extract or capture features/knowledge of correlations between RCT and ST. Since the GAN models have the aforementioned limitations, we devised a new translation method: Since DFT can extract features regardless of the location, a new translation method adopts DFT in this study. As a result, substitution learning is motivated by the knowledge transfer that exploits the correlations between ST and RCT, and DFT is employed due to its feature extraction process regardless of the target objects' locations. Therefore, the CSS-Net based on multi-task learning includes three pipelines; (1) as the main task, the segmentation task of RCT and ST regions. (2) the classification task for determining the presence or absence of RCTs, and (3) the substitution task based on DFT that substitutes ruptured ST (RCT) images with normal ST images.

To summarize, the main contributions are summarized as follows:

- **Substitution learning**: In terms of data augmentation, substitution learning achieves reliable data manipulation using DFT compared to GANs which utilize intensity-based feature maps.
- **Multi-task learning**: In terms of knowledge transfer, the CSS-Net improves the segmentation performance with the interactions between three modules of substitution, classification, and segmentation.

– **Diagnostic performance**: In terms of segmentation, the CSS-Net achieves 10% improved RCT diagnostic performance compared to the baseline model using proposed modules, and $6 \sim 8\%$ improved RCT diagnostic performance.

2 Methods

This section illustrates the detailed architecture of the CSS-Net and its design principle. First, we introduce the architecture of the CSS-Net with the individual pipelines for multi-task learning. Then, the detailed descriptions of the substitution learning in the CSS-Net follow. Table 1 summarizes the mathematical notation to construct the CSS-Net.

Table 1. Mathematical notations for the CSS-Net. Here, $[G^{seg}(C=c)]_{h,w} = 1$ *iff* $[G^{seg}]_{h,w,c} = 1$ else 0.

Notation	Dimension	Element	Related notations	Description
I	$\mathbb{R}^{H\times W}$	$[I]_{h,w} \in [0,1]$	H: height of I W: width of I	Input MRI image
$M_{seg}(I)$	$\mathbb{R}^{H\times W\times 3}$	$[M_{seg}(I)]_{h,w,c}$ $= p \in [0,1]$	$M_{seg}(I)\|_c \in \mathbb{R}^{H\times W}$ $M_{seg}(I)\|_{h,w} \in \mathbb{R}^{C}$	Prediction by segmentation module
G^{seg}	$\mathbb{R}^{H\times W\times 3}$	$[G^{seg}]_{h,w,c}$ $= g \in \{0,1\}$	$G^{seg}(C=c) \in \mathbb{R}^{H\times W}$	Ground truth in segmentation task
$M_{clf}(I)$	$\mathbb{R}^2$	$M_{clf}(I) = \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}$	$M_{clf}(I)\|_c = p_c \in [0,1]$ $\sum_c M_{clf}(I)\|_c = 1$	Prediction by classification module
G^{clf}	$\mathbb{R}^2$	$G^{clf} = \begin{pmatrix} p_0 \\ p_1 \end{pmatrix}$	$G^{clf}_c = p_c \in \{0,1\}$	Ground truth in classification task
DFT			$DFT : \mathbb{R}^{H\times W} \to \mathbb{R}^{H\times W}$	DFT function
$IDFT$			$IDFT : \mathbb{R}^{H\times W} \to \mathbb{R}^{H\times W}$	Inverse DFT function
S			$S : \mathbb{R}^{H\times W} \to \mathbb{R}^{H\times W}$	CNNs in M_{sl}
$M_{sl}(I)$	$\mathbb{R}^{H\times W}$	$[M_{sl}(I)]_{h,w} \in [0,1]$	$I' = IDFT(X')$ $X' = S(DFT(I * G(C=2)))$	Substituted image
CL			$CL : \mathbb{R}^{H\times W} \times \mathbb{R}^{H\times W} \to \mathbb{R}$	Content loss function

2.1 Multi-Task Learning Architecture

Fig. 2 (a) describes the overall architecture and pipeline of the CSS-Net, which includes the substitution (M_{sl}), classification (M_{clf}), and segmentation (M_{seg}) modules based on convolutional neural networks (CNNs). The CSS-Net aims to predict the multi-categorical segmentation masks of the background (BG), ST, and RCT. To this end, the CSS-Net is mainly designed for the segmentation task

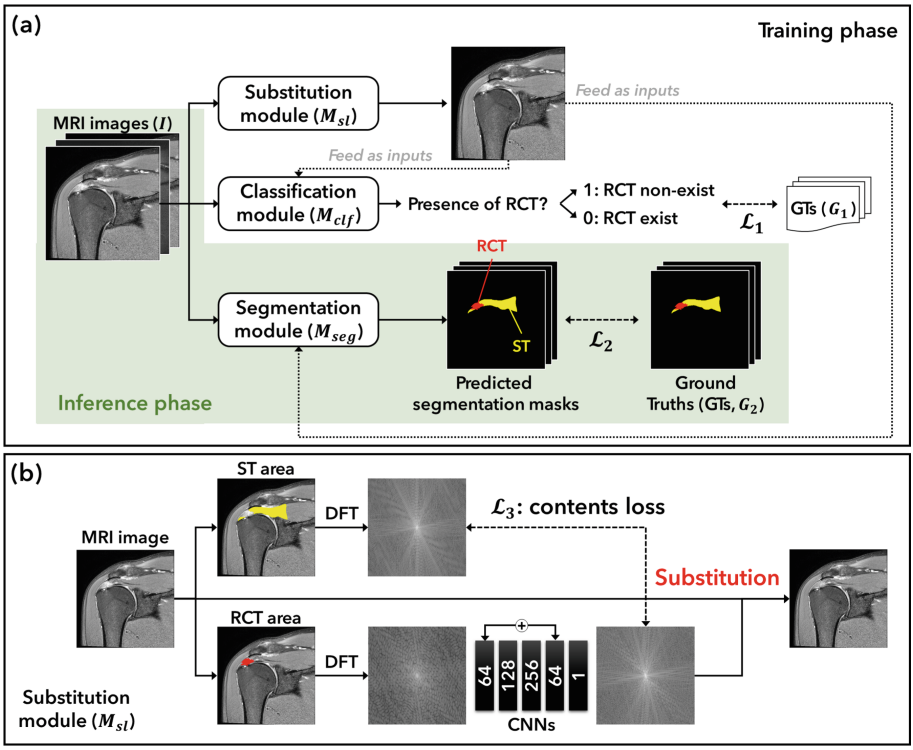

Fig. 2. (a) Overall pipeline of CSS-Net. Each module is optimized using the corresponding loss functions in the training phase. In the inference phase, the CSS-Net predicts the segmentation masks of ST and RCT using only the segmentation module. (b) Detailed architecture of the substitution learning module.

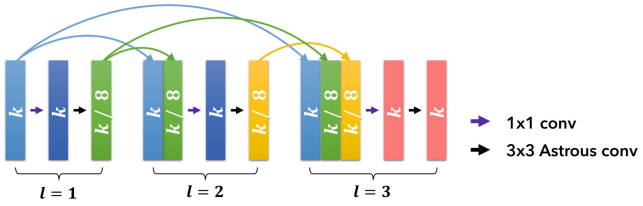

Fig. 3. Dense Block in the CSS-Net. Since ST and RCT occupy a small area, it is required to enlarge receptive fields to comprehend the correlations between RCT and ST. To this end, the Astros convolutions are utilized.

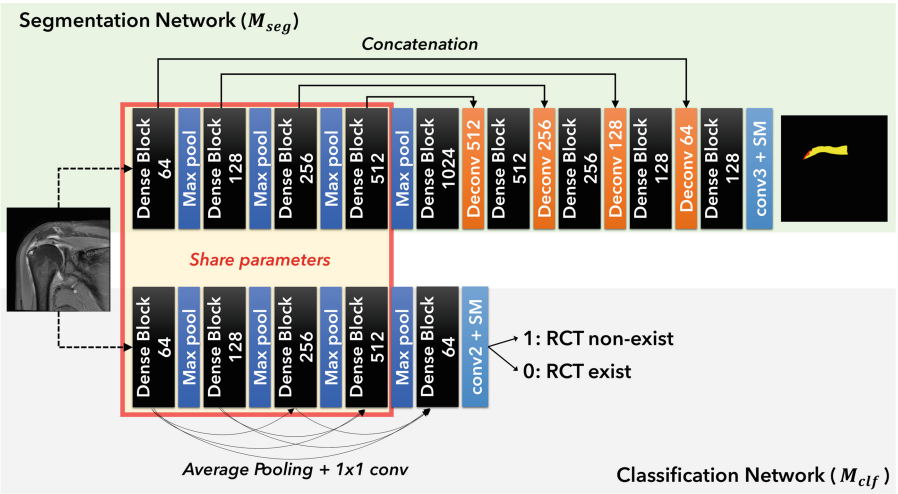

Fig. 4. Detailed architecture of M_{seg} and M_{clf} in the CSS-Net. M_{seg} and M_{clf} are designed based on U-Net and VGGNet, respectively. Convolution blocks in original models are replaced with DenseBlock.

(main task). In addition, despite the feasibility of single utilization of M_{seg}, to enhance the feature extraction during the optimization, two supplementary tasks and modules are appended; the classification module (M_{clf}) and the substitution learning module (M_{sl}). Figures 3 and 4 illustrate the detailed architecture of the CSS-Net. Note that, several convolutions are shared between M_{seg} and M_{clf} to transfer the learned knowledge related to RCT as illustrated in Fig. 4.

2.2 Segmentation Task

As the main task, M_{seg} aims to generate one-hot labeled segmentation masks that include three categories of $\{0, 1, 2\}$, where 0, 1, and 2 indicate BG, ST, and RCT classes, respectively. M_{seg} is constructed based on the U-Net [24], Atrous convolution in DeepLab [25], and the dense connectivity [26], which is a CNN structure shared with M_{clf} (Fig. 3). The segmentation network M_{seg} extracts the features of the input MRI images ($I \in \mathbb{R}^{H \times W}$), and then generates the segmentation outputs ($M_{seg}(I) \in \mathbb{R}^{H \times W \times 3}$). Here, $M_{seg}(I)$ is the probability-based segmentation mask, and thus $\sum_{c}^{\{0,1,2\}} M_{seg}(I)|_c = \mathbf{1} \in \mathbb{R}^{H \times W}$, where c indicates class-wise notation. In addition, the pixel at the location (h, w) is classified as $\underset{c}{\text{argmax}}\big(M_{seg}(I)|_{h,w}\big)$. Note that M_{seg} is trained with M_{clf} and M_{sl} during training, but the single M_{seg} is used during inference.

2.3 Classification Task

As the supplementary task, M_{clf} classifies the existence of the RCT in I as a binary classification, wherein the classification category is 0 or 1, where 0

and 1 indicate the absence and presence of RCT in I. Here, M_{clf} is designed using the dense connectivity [26] and shares parameters with M_{seg}. The M_{seg} extracts features of I and then outputs the classification results ($M_{clf}(I) \in \mathbb{R}^2$). Since $M_{clf}(I)$ is also probability-based matrix, $\sum_c^{\{0,1\}} M_{clf}(I)|_c = 1 \in \mathbb{R}$, and is determined as $\underset{c}{\text{argmax}}\big(M_{clf}(I)\big)$. M_{clf} transfers the learned knowledge about the RCT by sharing parameters between M_{seg} and M_{clf}, and this knowledge improves the performance of M_{seg}.

2.4 Substitution Learning

M_{sl} substitutes the RCT region, which is a ruptured ST area in the MRI images, for a normal ST style. The substituted images are then utilized as additional inputs for M_{seg} and M_{clf}, in terms of data augmentation. Figure 2(b) describes the detailed pipeline of M_{sl}. First, I is binary-masked using the corresponding ground truth ($G = G^{seg}$) with the two outputs ($I{*}G(c=1)$) and RCT ($I{*}G(c=2)$), where $*$ indicates the Hadamard product. The individual masked regions are then converted into the frequency-domain as shown in Figs. 2(b) and 5(a). Here, the DFT is formulated as follows:

$$F[x,y] = \frac{1}{HW}\sum_{h}^{H}\sum_{w}^{W} I[h,w]e^{-j2\pi\left(\frac{h}{H}y+\frac{w}{W}x\right)}, \quad j=\sqrt{-1} \tag{1}$$

where, F is the output mapped into the frequency domain, e is Euler's number, and H and W are the height and width of I, respectively. Subsequently, a simple CNN architecture (S) with identical mapping transfers the DFT-converted output of the RCT, which is $D = DFT(I{*}G(c=2))$, as $D' = (S{\circ}DFT)(I{*}G(c=2))$. The inverse DFT (IDFT) is applied to D', and the substituted images $IDFT(D')$ is finally generated. In summary, the substituted image (I') is calculated as $I' = M_{sl}(I) = (IDFT \circ S \circ DFT)(I * G(c=2))$. As illustrated in Fig. 5(b), the generated images by SL are more reliable than those of GANs.

2.5 Loss Functions of CSS-Net

As illustrated in Fig. 2, the CSS-Net includes three loss functions of classification loss ($\mathcal{L}_1$), segmentation loss ($\mathcal{L}_2$), and substitution loss ($\mathcal{L}_3$). Here, $\mathcal{L}_1$ and $\mathcal{L}_2$ are based on the cross-entropy loss function. In particular, the KL-divergence of $M_{clf}(I)|_c$ is compared to that of the corresponding ground truth G^{clf}, and thus $\mathcal{L}_1 = \sum_c G_c^{clf} \log M_{clf}(I)|c$. Likewise, $\mathcal{L}_2 = \sum_{h,w,c}[G^{seg}]_{h,w,c} \log[M_{seg}(I)]_{h,w,c}$. Additionally, the content loss (CL) is utilized to compare the similarity between the substituted RCT and the normal ST, which is regarded as ground truth, in the frequency-domain. In particular, $(S \circ DFT)(I * G(C=2))$ is compared to $I{*}G(C=1)$, and thus $\mathcal{L}_3 = CL(S(DFT(I{*}G(C=2))), I{*}G(C=1))$. Moreover, the CSS-Net has additional constraints if the RCT does not exist in I, then the substituted images are the same as the original image, and thus $I = M_{sl}(I)$. Therefore, $G_0^{clf}|I - M_{sl}(I)|_1$ is constrained, where $f(x) = |x|_1$ is l_1 loss. Besides, since the RCT is not in substituted image, $M_{clf}(M_{sl}(I))|_0 = 1$ is constrained (Table 2).

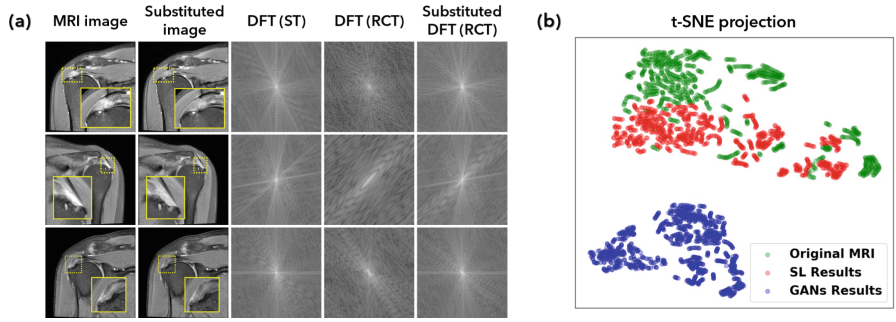

Fig. 5. (a) MRI, substituted, and corresponding DFT images. (b) t-SNE projection of MRI (green), SL images (red), and GANs images (blue). Distribution of substituted images is more similar to MRI than that of medical style GANs [27] (Color figure online).

Table 2. Loss functions for training CSS-Net. $\mathcal{L}_{total} = \sum_{i=1}^{5} \alpha_i \mathcal{L}_i$. Here α_i is a scale factor and trainable. The initial values of α_i is 1.0 except for $\alpha_4 = 0.01$. *CE* and *CL* are cross-entropy and content loss.

Loss function	Definition	Description
$\mathcal{L}_1$	$\sum_c G_c^{clf} \log M_{clf}(I)\|c$	CE for classification task
$\mathcal{L}_2$	$\sum_{h,w,c}[G^{seg}]_{h,w,c} \log[M_{seg}(I)]_{h,w,c}$	CE for segmentation task
$\mathcal{L}_3$	$CL(S(DFT(I * G(C = 2))), I * G(C = 1))$	CL for substitution task
$\mathcal{L}_4$	$G_0^{clf}\|I - M_{sl}(I)\|_1$, and $\|x\|_1$ is l_1 loss	If RCT not in I, then $I = M_{sl}(I)$
$\mathcal{L}_5$	$1 - M_{clf}(M_{sl}(I))\|_0$	RCT is not in $M_{sl}(I)$

3 Experiments and Results

3.1 Dataset Construction and Environmental Set-up

The data collection has been conducted in accordance with the Declaration of Helsinki, the protocol was approved by the Ethics Committee of the Institutional Review Board of Daegu Catholic University Medical Center, and the clinical

Table 3. Total number of samples and ratio in each segmentation class.

	Total	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Avg. # pixels per images (%)	
Total patients	42	9	9	8	8	8		
Total images	889	196	192	166	174	161		
BG	612	127	131	119	128	109	Background (BG)	99.20%
BG + ST	123	31	33	13	17	29	Supraspinatus tendon (ST)	0.76%
BG + ST + RCT	152	38	28	34	29	23	Rotator cuff tear (RCT)	0.04%

Table 4. Training environment for the CSS-Net.

Parameter	Value	Parameter	Value	Parameter	Value
Image Size	512	Resolution	16bits	Augmentation	Flip
Optimizer	Adam	Learning rate	1e-3	Batch size	16
β_1 in Adam	0.9	β_2 in Adam	0.99	ϵ in Adam	1e-7
CPUs	2 Xeons	GPUs	8 Titan-Xps	RAM	256GB
Layer	Value			Layer	Value
Normalization	Group Normalization ($G = 16$)			Activation	ReLU

trial in this paper has been in accordance with ethical standards. In total, 889 images were obtained from 42 patients with shoulder pains. Table 3 illustrates the detailed description of the dataset. In the acquired dataset, the number of images on ST is 123, and that on both the ST and RCT is 152. The other images did not include the ST or RCT. The acquired dataset were divided into 5 folds for the k-fold cross-validation to guarantee the robustness of the experiments, such that each fold contained at least 160 images. MRI images originating from a single patient were only included in single folds. In addition, the experimental environment and the hyper-parameters to train deep learning models are illustrated in Table 4.

Table 5. Comparisons between ours with state-of-the-art models.

	mIoU	IoU-BG	IoU-ST	IoU-RCT	Sensitivity-RCT
U-Net	0.65 ± 0.01	0.93 ± 0.01	0.62 ± 0.02	0.38 ± 0.05	0.68 ± 0.02
DeepLabV3+	0.69 ± 0.01	0.94 ± 0.02	0.66 ± 0.03	0.41 ± 0.04	0.71 ± 0.02
IteR-MRL	0.71 ± 0.01	0.97 ± 0.01	0.69 ± 0.02	0.43 ± 0.06	0.80 ± 0.03
SA-RE-DAE	0.72 ± 0.01	0.97 ± 0.01	0.72 ± 0.02	0.45 ± 0.05	0.78 ± 0.02
Seg + Clf + SL	**0.75 ± 0.01**	**0.98 ± 0.01**	**0.74 ± 0.02**	**0.51 ± 0.03**	**0.88 ± 0.04**

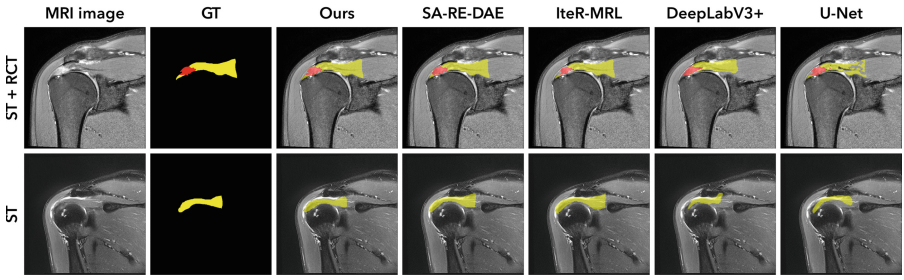

Fig. 6. Representative segmentation results using deep learning models.

3.2 Comparison with State-of-the-Art Models

Table 5 illustrates the quantitative analysis of the proposed CSS-Net compared with other deep learning models. U-Net and DeepLabV3+ were employed because of their popularity in segmentation tasks. In addition, the SA-RE-DAE [28] and IteR-MRL [29] were utilized as state-of-the-art segmentation and multi-task models. The experimental results demonstrated that all models achieve high scores in segmenting BG. Expecting the RCT, the CSS-Net significantly outperforms the other models. It showed at least a 6% IoU-RCT compared to the other models. In particular, the CSS-Net achieved 10% $\sim$ 20% improved sensitivity in RCT segmentation, suggesting that the CSS-Net with substitution learning could be utilized as an excellent diagnostic tool to localize RCT, as shown in Fig. 6.

3.3 Analysis of Our Model

Since the CSS-Net was designed based on multi-task learning that includes segmentation, classification, and substitution learning, an ablation study was conducted by using each task (Seg, Seg + Clf, Seg + SL, and Seg + Clf + SL). Additionally, because substitution learning was comparable with GANs, CSS-Net, which replaced the substitution module with Style-GAN [27] was also compared. Table 6 illustrates the ablation study of the CSS-Net. The results exhibited that the multi-task learning of segmentation and classification could slightly improve the segmentation of the RCT (Seg and Seg + Clf). In contrast, generative tasks, including GAN and SL, could significantly improve the performance of the CSS-Net. Here, the CSS-Net with SL and Clf tasks improve at 8% IoU-RCT than the baseline. However, the SL-based generative task was preferred in teaching intensity- and frequency-domain knowledge rather than GAN-based style transfer. The results implied that informative knowledge in SL could be transferred into the Seg task.

Table 6. Ablation study of CSS-Net.

	mIoU	IoU-BG	IoU-ST	IoU-RCT	Sensitivity-RCT
Seg (baseline)	0.68 ± 0.01	0.96 ± 0.01	0.65 ± 0.03	0.41 ± 0.04	0.69 ± 0.02
Seg + Clf	0.69 ± 0.01	0.95 ± 0.01	0.67 ± 0.03	0.44 ± 0.05	0.74 ± 0.02
Seg + SL	0.71 ± 0.01	0.96 ± 0.01	0.70 ± 0.04	0.46 ± 0.03	0.78 ± 0.02
Seg + GAN	0.71 ± 0.03	0.96 ± 0.02	0.70 ± 0.06	0.47 ± 0.11	0.75 ± 0.05
Seg + Clf + GAN	0.73 ± 0.02	0.97 ± 0.02	0.72 ± 0.05	0.49 ± 0.09	0.81 ± 0.05
Seg + Clf + SL	**0.75 ± 0.01**	**0.98 ± 0.01**	**0.74 ± 0.02**	**0.51 ± 0.03**	**0.88 ± 0.04**

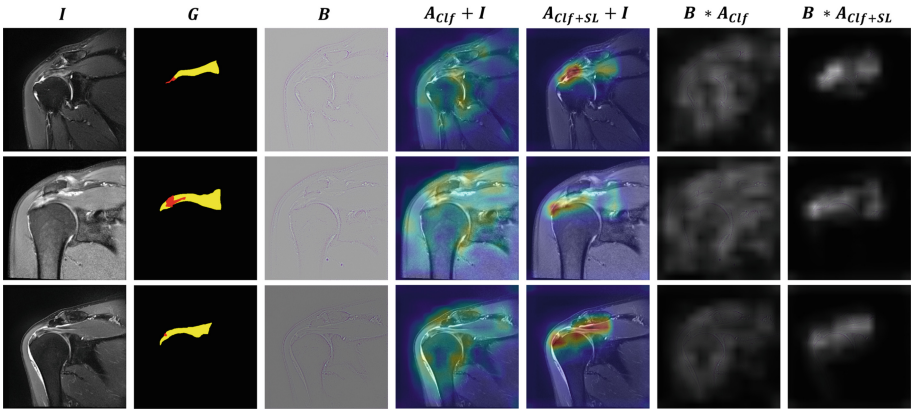

Fig. 7. Representative results of Guided Grad-CAMs. Left→Right: MRI images (I), Ground truth (G), Guided-backprop (B), Overlay of I and Grad-CAM (A_{Clf}) by the *Clf*-network which has only classification module, Overlay of I and Grad-CAM (A_{Clf+SL}) by the *Clf+SL*-network which has classification and substitution module, Guided Grad-CAM ($B{*}A_{Clf}$) by the *Clf*-network, and Guided Grad-CAM ($B{*}A_{Clf+SL}$) by the *Clf+SL*-network.

4 Discussion and Future Work

4.1 Explainability

To analyze the effectiveness of substitution learning on other tasks, we compared the Grad-CAM [12] of the CSS-Net with and without an SL module in the RCT classification. Figure 7 illustrates the Grad-CAM and Guided Grad-CAM samples. The results demonstrated that the Grad-CAMs generated by the CSS-Net without the SL module ($A_{clf} + I$) widely exhibited attentions nearby shoulders. On the contrary, the Grad-CAMs generated by CSS-Net with the SL module ($A_{clf+sl} + I$) exhibited an integrated attention distribution similar to that by ground truth of ST (G). The results implied that the CSS-Net with the SL module extracted the features maps of RCT from the ST- and RCT-related areas rather than the entire image. Therefore, it was concluded that the substitution learning improved RCT-related feature extraction by learning the correlations between ST and RCT.

4.2 Limitations and Improvements

One of the main reasons for low IoU-RCT values was the imbalanced pixel distribution in the dataset. Since the BG pixels occupied approximately 99%, whereas the RCT pixels occupy 0.04%, misprediction of the BG significantly affected the

accuracy of the RCT regions. Although substitution learning improved segmentation performance by reliable data manipulation than GANs and by transferring the informative knowledge into the segmentation modules, the imbalanced problem could be further improved. Additionally, since the multi-task deep learning models demanded heavy memory utilization owing to their large number of CNNs, a long training time and high memory cost are required to optimize the deep learning models. However, we improved the prediction time by eliminating other tasks except for the segmentation task in the prediction phase. Therefore, the CSS-Net costs the same prediction time as the baseline but it offered high performance in segmentation. Furthermore, we have mainly focused on segmenting the ST and the RCT. However, substitute learning could be further extended to diagnose any diseases that are significantly imbalanced by learning the correlations between normal and disease regions using substitution (abnormal to normal).

5 Conclusions

We introduced integrated multi-task learning as an end-to-end network architecture for RCT segmentation in MRI images. We also proposed a novel substitution learning using DFT to augment data more reliably for the imbalanced dataset, as well as to improve accuracy by knowledge transfer. We employed the SL instead of GANs-based approaches since the SL was demonstrated as more reliable than GANs with even low computation costs. Our results showed that the CSS-Net produced a superior segmentation performance owing to the abundant knowledge transfer from the classification and substitution tasks to the segmentation task, outperforming other state-of-the-art models. It showed a 10% higher IoU value than the baseline model, and even at least 6% higher IoU values than those shown by other state-of-the-art models. Further experiments should be conducted for clinical applications that require reliable data augmentation and high performance.

Acknowledgments. This work was partially supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: RS-2022-00141185). Additionally, this work was supported by the Technology Innovation Program(2001424, Development of an elderly-friendly wearable smart healthcare system and service for real-time quantitative monitoring of urination and defecation disorders) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea). Ground truths were generated by one MRI imaging specialist and one of lab members, and reviewed by two orthopedic surgeons. The code and the dataset are released in the public repository[1]([1] https://github.com/kyungsu-lee-ksl/ACCV2022-Substitution-Learning.git).

References

1. Moosikasuwan, J.B., Miller, T.T., Burke, B.J.: Rotator cuff tears: clinical, radiographic, and us findings. Radiographics **25**, 1591–1607 (2005)
2. Davidson, J.J., Burkhart, S.S., Richards, D.P., Campbell, S.E.: Use of preoperative magnetic resonance imaging to predict rotator cuff tear pattern and method of repair. Arthrosc. J. Arthroscopic Relat. Surg. **21**, 1428–e1 (2005)
3. Shin, Y.K., Ryu, K.N., Park, J.S., Jin, W., Park, S.Y., Yoon, Y.C.: Predictive factors of retear in patients with repaired rotator cuff tear on shoulder MRI. Am. J. Roentgenol. **210**, 134–141 (2018)
4. Kukkonen, J., Kauko, T., Virolainen, P., Äärimaa, V.: The effect of tear size on the treatment outcome of operatively treated rotator cuff tears. Knee Surg. Sports Traumatol. Arthrosc. **23**, 567–572 (2015)
5. Khazzam, M., et al.: Magnetic resonance imaging identification of rotator cuff retears after repair: interobserver and intraobserver agreement. Am. J. sports Med. **40**, 1722–1727 (2012)
6. Medina, G., Buckless, C.G., Thomasson, E., Oh, L.S., Torriani, M.: Deep learning method for segmentation of rotator cuff muscles on MR images. Skeletal Radiol. **50**, 683–692 (2021)
7. Mazurowski, M.A., Buda, M., Saha, A., Bashir, M.R.: Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. J. Magn. Reson. Imaging **49**, 939–954 (2019)
8. Roth, H.R., et al.: Federated whole prostate segmentation in MRI with personalized neural architectures. In: de Bruijne, M. (ed.) MICCAI 2021. LNCS, vol. 12903, pp. 357–366. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_34
9. Astuto, B., et al.: Automatic deep learning-assisted detection and grading of abnormalities in knee MRI studies. Radiol. Artif. Intell. **3**, e200165 (2021)
10. Kim, M., Park, H.m., Kim, J.Y., Kim, S.H., Hoeke, S., De Neve, W.: MRI-based diagnosis of rotator cuff tears using deep learning and weighted linear combinations. In: Doshi-Velez, F. (eds.) Proceedings of the 5th Machine Learning for Healthcare Conference. Vol. 126 of Proceedings of Machine Learning Research, pp. 292–308. PMLR (2020)
11. Shim, E., et al.: Automated rotator cuff tear classification using 3D convolutional neural network. Sci. Rep. **10**, 1–9 (2020)
12. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)
13. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit. Imaging **32**, 582–596 (2019)
14. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
15. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
16. Lee, K., Kim, J.Y., Lee, M.H., Choi, C.H., Hwang, J.Y.: Imbalanced loss-integrated deep-learning-based ultrasound image analysis for diagnosis of rotator-cuff tear. Sensors **21**, 2214 (2021)

17. Northcutt, C., Jiang, L., Chuang, I.: Confident learning: estimating uncertainty in dataset labels. J. Artif. Intell. Res. **70**, 1373–1411 (2021)
18. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. **27** (2014)
19. Antoniou, A., Storkey, A., Edwards, H.: Data augmentation generative adversarial networks. arXiv preprint arXiv:1711.04340 (2017)
20. Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., Malossi, C.: BAGAN: data augmentation with balancing GAN. arXiv preprint arXiv:1803.09655 (2018)
21. Shin, H.-C., et al.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Gooya, A., Goksel, O., Oguz, I., Burgos, N. (eds.) SASHIMI 2018. LNCS, vol. 11037, pp. 1–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00536-8_1
22. Calimeri, F., Marzullo, A., Stamile, C., Terracina, G.: Biomedical data augmentation using generative adversarial neural networks. In: Lintas, A., Rovetta, S., Verschure, P.F.M.J., Villa, A.E.P. (eds.) ICANN 2017. LNCS, vol. 10614, pp. 626–634. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68612-7_71
23. Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M.: Data augmentation using generative adversarial networks (cycleGAN) to improve generalizability in CT segmentation tasks. Sci. Rep. **9**, 1–9 (2019)
24. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
25. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
26. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)
27. Kazuhiro, K., et al.: Generative adversarial networks for the creation of realistic artificial brain magnetic resonance images. Tomography **4**, 159–163 (2018)
28. Khan, S.H., Khan, A., Lee, Y.S., Hassan, M., et al.: Segmentation of shoulder muscle MRI using a new region and edge based deep auto-encoder. arXiv preprint arXiv:2108.11720 (2021)
29. Liao, X., et al.: Iteratively-refined interactive 3D medical image segmentation with multi-agent reinforcement learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9394–9402 (2020)