JCSE Journal of Computing Science and Engineering

Regular Paper

Journal of Computing Science and Engineering,
Vol. 16, No. 1, March 2022, pp. 43-51

Domain Transformation for PMMA-Inserted Vertebral Body Segmentation

Minyoung Park and Jinah Park*
School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Korea
my.park@kaist.ac.kr, jinah@cs.kaist.ac.kr

Abstract
Learning-based medical image segmentation has been advanced with the collection of datasets and various training methodologies. In this work, we target bone cement (polymethylmethacrylate [PMMA]) inserted vertebral body segmentation, where the target dataset was relatively scarce, compared to a large-scale dataset for the regular vertebra segmentation task. We presented a novel domain transformation framework, where a large-scale training set for our target task was generated from the existing dataset of a different domain. We proposed two main components: label translation and image translation. Label translation was proposed to filter out unnecessary regions in a segmentation map for our target task. In addition to label translation, image translation was proposed to virtually generate PMMA-inserted images from the original data. The synthesized training set by our method successfully simulated the data distribution of the target task; therefore a clear performance improvement was observed by the dataset. By extensive experiments, we showed that our method outperformed baseline methods in terms of segmentation performance. In addition, a more accurate shape and volume of a bone were measured by our method, which satisfied the medical purpose of segmentation.

Category: Smart and Intelligent Computing

Keywords: Computer vision; Medical image segmentation; Domain transformation

I. INTRODUCTION

Learning-based medical image segmentation has achieved state-of-the-art performance in various medical fields [1-3]. Such methods require a large dataset that consists of image and segmentation label pairs. Due to the efforts of many researchers, there are several large open datasets for each medical region of interest, e.g., brain [4] and vertebrae [5]. However, when it comes to real-world data, domain shift exists from large open datasets depending on the target task. To be specific, real-world target data may contain artifacts or lesions according to the purpose of segmentation, while a large open dataset may not contain. In addition, the segmentation label of a large open dataset may be related but may not perfectly match the target task, e.g., the target segmentation area is a subset of a labeled one. In this paper, a novel domain transformation framework was proposed that dealt with domain shift in both image and label, so that we could effectively utilize a large open dataset.

In this paper, the segmentation of a bone cement inserted vertebral body was targeted by learning-based segmentation. Vertebral body fractures caused by osteoporosis or bone tumor, bring on significant back pain [6]. The percutaneous vertebroplasty surgery alleviates the pain by bone cement injection; the bone cement is

Open Access http://dx.doi.org/10.5626/JCSE.2022.16.1.43 http://jcse.kiise.org

Received 14 December 2021; **Accepted** 02 March 2022
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

injected into fractured vertebral bodies, which increases the strength and stability of the bone. After the surgery, medical doctors need to measure the accurate shape and volume of the vertebral body for analyzing procedure-related morphologic patterns and the risk factors for the complications [7]. So, a precise vertebral body area is needed to be segmented. Since the bone cement often causes occlusions and unusual intensity values, a classical segmentation method tends to fail, thus a learning-based method is required for this task.

Building a learning-based segmentation model for the target task is confronted by the domain shift between the target and an open dataset. To train the segmentation model, we needed a large number of bone cement inserted vertebral body image and label pairs. However, manually collecting a large dataset is difficult since the number of medically treated subjects is limited, and segmenting a size of approximately 10^4 voxels each is laborious. Rather than collecting and labeling on the target domain, one might consider utilizing an off-the-shelf large open dataset for vertebra segmentation; it is related to our task but has a domain shift in terms of image and label. As expected, training a model on the original dataset showed poor performance on the target task (in our experimental result). In this paper, we filled the shift by building a rendering pipeline, which generates target-like data samples from the shifted data samples. As a result, we can utilize a large open dataset effectively by the domain transformation to bone cement inserted vertebral body segmentation.

Our contributions can be summarized as follows. First, a domain transformation method is introduced that includes both label translation and image translation. Second, we demonstrate that the proposed domain transformation makes segmentation performance improve in various experimental settings. Finally, experimental results showed that the methods help a model infer more precise volume and predict a more accurate 3D shape of bone-cement inserted vertebral body than other existing methods which satisfied the clinical purpose of the segmentation.

II. RELATED WORK

Medical image segmentation has been actively studied, since it is a fundamental task in medical image analysis and computer-aided diagnosis. For instance, medical image segmentation has been studied in terms of liver segmentation [8], brain segmentation [9], cell segmentation [10] and lung segmentation [11], amongst others.

A. PMMA-Inserted Vertebral Body Segmentation

In this paper, focus was on the polymethylmethacrylate (PMMA)-inserted vertebral body segmentation, where the region of interest, vertebral body, has been treated by an injection surgery with a bone-cement called PMMA; the injection into a fractured vertebral body bone increased the strength and stability of the vertebral body. After the surgery, the volume and the shape of a PMMA-inserted vertebral body was needed to be measured accurately to analyze structural changes of the treated vertebral body and clarify the risk factors of complications [7]. The segmentation task inferred the region of the vertebral body, from which the shape and the volume can be measured accordingly.

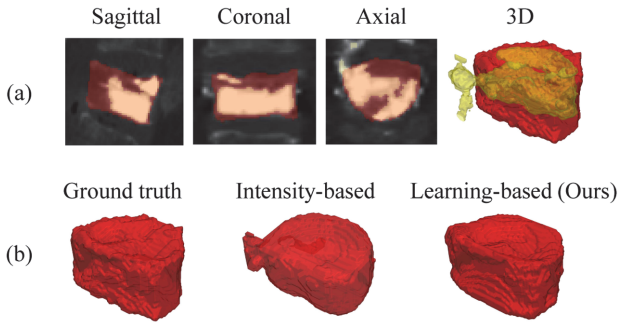

Fig. 1. (a) PMMA-inserted vertebral body in CT image with multi-view. The red area indicates the vertebral body and the yellow area indicates inserted PMMA, bone cement. A 3D rendered view shows that some inside and boundary of the vertebral body is occluded by PMMA, which makes PMMA-inserted vertebral body segmentation challenging. (b) Vertebral body ground truth and segmentation results. The introduction of artifacts, such as PMMA, gives confusion to the intensity-based approach. (Intensity-based segmentation is done by 3D graph cut. Density function from training data larger than 0.5 is given as segmented object seed, and density function zero is given as background seed. Code is from https://github.com/mjirik/imcut.) On the contrary, our learning-based approach showed a robust prediction result.

PMMA posed a challenge to the segmentation, since the region of interest, i.e., vertebral body, was often overlapped by the material, which made the task difficult due to the occluded area. As shown in Fig. 1(a), inserted PMMA leakage from vertebral body often occurs [12], so that PMMA penetrated the vertebral body boundary; PMMA was observed with high intensity and opacity in the computed tomography (CT) image, and some inside areas and boundaries of the vertebral body were occluded by PMMA. These occlusions made the true regions hidden behind and these required a solver to infer about the region. In this work, our approach used a data-driven and learning-based method to learn about obscure regions from data.

B. Vertebra and Vertebral Body Segmentation

Traditionally, vertebral body segmentation has been solved by intensity-based methods. Aslan et al. [13] solved the vertebral body segmentation in CT image by graph cuts using given 3D shape prior knowledge. Hille et al.

[14] solved vertebral body segmentation in magnetic resonance imaging (MRI) with a hybrid level-set based approach. However, the presence of artifacts like PMMA renders intensity-based approaches ineffective for our purpose; these methods assume that voxels with the same label have equal intensities. However, in our task, intensity often gives misleading information; a PMMA region showed distinctively higher intensity and possibly leads to misclassification. The segmentation result of graph cuts with 3D shape prior, i.e., intensity-based technique, as shown in Fig. 1(b), confused PMMA boundary with vertebral body boundary, resulting in poor vertebral body prediction. Furthermore, as the approaches were based on intensity-based features, they were inappropriate to predict the boundary of an occluded object.

Our approach was to exploit a large-scale dataset to make our model robust to those challenges. A learning-based vertebral body segmentation method [15] solved vertebral body segmentation by combining estimated likelihood and prior probability, which were learned from training data. Closely related to our findings, vertebra segmentation has been studied extensively due to the large open dataset, VerSe [5], which was introduced for a challenge. With the large open dataset, VerSe, vertebra segmentation has achieved a high dice score of 0.94 [2] based on U-Net [3, 10]. In our work, the vertebra dataset was exploited—which was in a different domain from our task—by transforming its images and labels to fit our task.

Collecting and labeling specific tomographies of PMMA-inserted vertebra was difficult since the number of medically treated subjects was limited and manually labeling 3D images was laborious. That is why several learning-based vertebra or vertebral body segmentation work has focused on non-occluded cases, e.g., vertebral body [15] and vertebra [5]. To ameliorate this, we proposed to learn with a small target dataset by generating large-scale training data for our task with the existing non-occluded dataset of vertebrae.

C. Domain Transformation

Generating training data by transforming source data for a target task has been studied [16, 17]. The generation process was often performed via the transformation from source domain with easily obtainable, to a target domain where data is expensive. For instance, Sandfort et al. [18] generate non-contrast CT from contrast CT data, where public CT dataset consisted of contrast CT although real-world data contain a few non-contrast CT. Veit et al. [19] generated clean labels from the noisy labels, annotated with noise, where noisy labels were easily obtainable by image classifier and clean labels were from costly human verification. In this way, prior works of domain transformation included image translation or label translation. Our framework differs from prior works in that the domains transformed in both image and label, so as to utilize the source dataset even if domain shift existed in both image and label at the same time.

In this study, we propose a novel domain transformation framework translating large-scale training data to our target data. This domain modification was used to increase performance on our task, PMMA-inserted vertebral body segmentation.

III. METHODS

In this section, we introduce our domain transformation framework for PMMA-inserted vertebral body segmentation, summarized in the graphical description shown in Fig. 2. To recap, our goal is to segment the vertebral body from CT images having PMMA. To achieve this goal, our approach is to generate target domain data from relatively easy-to-obtain source data.

In the perspective of machine learning, the data was divided into labels and images; we transformed both to achieve our goal. Formally, we have an easily-obtainable large dataset V comprising tuples of vertebra label

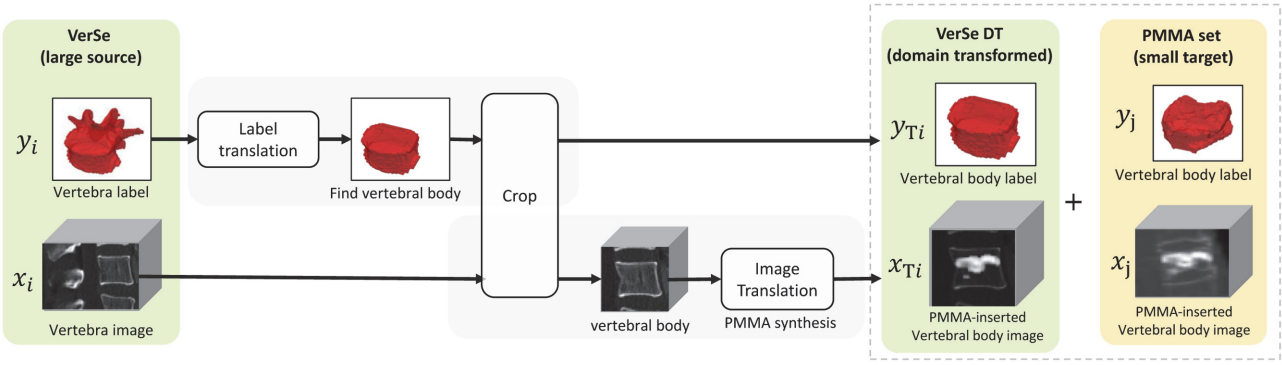

Fig. 2. Overview of the proposed domain transformation framework. Large source dataset, VerSe, is domain transformed to VerSe DT by image and label translation. The domain transformation matches the domain of VerSe and the small target dataset, PMMA set. Finally, making use of VerSe DT leads to performance improvement of PMMA-inserted vertebral body segmentation.

$y \in [0, 1]^{H \times W \times D}$ and vertebra CT image $x \in \mathbb{R}^{H \times W \times D}$, $V = \{(y_i, x_i), \ldots\}$, and a small target dataset P of tuples of vertebral body label $y \in [0, 1]^{H' \times W' \times D'}$ and PMMA-inserted vertebral body CT image $x \in \mathbb{R}^{H' \times W' \times D'}$, $P = \{(y_j, x_j), \ldots\}$. The two sets differed significantly in size with $|V| \gg |P|$. For instance, in our experiments, V exceeded P by the factor of 20. Since the labels and the images in V have the domain shift regarding our target task, our goal is to design an effective approach to transform the domain of V to match distribution with P. To realize this, we designed a function for domain transform:

$$(y_{Ti}, x_{Ti}) = \texttt{DomainTransform}((y_i, x_i)), \quad (1)$$

where $y_T \in [0, 1]^{H \times W \times D}$ and $x_T \in \mathbb{R}^{H \times W \times D}$. Note that $(y_i, x_i) \in V$ was domain transformed, so that we can synthesize a domain transformed large target dataset $T = \{(y_{Ti}, x_{Ti}), \ldots\}$. Our domain transformation contained two main components. First, a source label was translated to a target label by *label translation*. With a domain-specific design, our method extracted a vertebral body region from a vertebra label by a graph cuts algorithm (Section III-A). Second, an image in the source data did not include PMMA-like artifacts. To match image distribution, PMMA was synthesized to source picture by *image translation* (Section III-B). Finally, we introduce the loss functions that our segmentation model uses (Section III-C).

A. Label Translation

The translation of vertebra labels in V into vertebral body labels was required to transform vertebra label y_i sampled from V, so that y_i followed the distribution of labels in P. The proposed algorithm is depicted in Fig. 3 and detailed in Algorithm 1.

Algorithm 1 Label translation

Input: y such that $(y, x) \in V$ ▷ Vertebra label
Output: y_T such that $(y_T, x_T) \in T$ ▷ Vertebral body label
1: $M[j,k] = y[\frac{H}{2}, j, k]$ ▷ Mid-sagittal slice
2: $C = \texttt{FindContours}(M, threshold)$
3: $VB, VP = \texttt{FillContours}(C)$ ▷ Find seeds
4: $S =_S E(y, S, VB, VP)$ ▷ Graph cuts
5: $y_T =$ segmented object of S

Initializing seed for vertebral body: To automatically translate a vertebra label to a vertebral body label, we use shape prior like previous works [13, 20] since the shapes of bones have common features. Given an input label $y \in [0, 1]^{H \times W \times D}$ is in canonical voxel orientation; the axes are left to right, posterior to anterior, and inferior to superior. In line 1 of Algorithm 1, we pick $M \in [0, 1]^{W \times D}$, the mid-sagittal slice of y, where vertebral body and extra regions, i.e., vertebral process, are separated. Next, we find contours from M with contour area size larger than the threshold to eliminate contours by noise (line 2), followed by each contour in C being filled. We use the positional prior, i.e., the order of each bone, to determine the initial seed for vertebral processes VP and vertebral body VB (line 3); the smallest x, y coordinate in each contour is used for the process. In addition, the intersection areas of $VB\&VP$ along sagittal slices of y are also given as seeds to make a robust initial seed in 3-dimension.

Graph cuts to find vertebral body: The next step is to find the vertebral body by graph cuts. Graph cuts is a traditional computer vision technique to split two areas with given seeds. In the graph cuts, every voxel of vertebra label y is regarded as a node of a graph. In line 4, graph cuts algorithm finds a segmentation that minimizes the energy function E with the given seeds. By graph cuts, we find the smallest area to divide the vertebral body and extra regions. Finally, the segmented vertebral body is used as y_T, a domain translated vertebral body label (line 5).

After the label translation, we crop the label and image based on the translated vertebral body label; we put the label at the center, and then crop the image and label with margin, simultaneously. As illustrated in Fig. 2, the crop is another step to match the image and label distribution to the target dataset since the crop based on the translated label makes the image focus on the region of interest.

B. Image Translation

Image translation with PMMA synthesis aimed to match the distribution of images in V and P; original images from the large source dataset V did not include artifacts such as PMMA, while the target dataset P contained PMMA-inserted vertebral body images. In CT images, PMMA was shown in high intensity and sometimes leaking out of the vertebral body. The shape of PMMA was complicated,

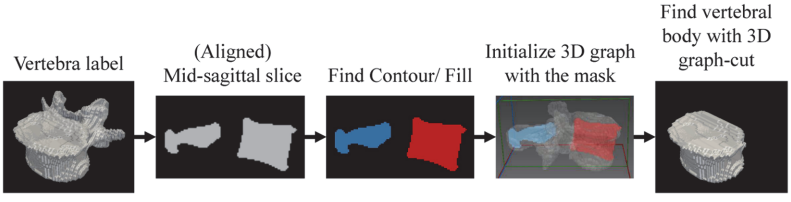

Fig. 3. Illustration of label translation. Red indicates vertebral body and blue indicates the extra region, i.e., vertebral process.

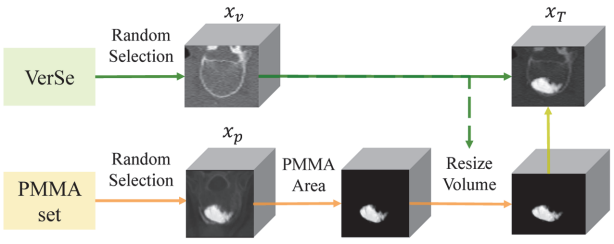

Fig. 4. Image translation.

due to the underlying anatomy and surgical technique. Thus, it was difficult to design a reasonable PMMA directly. Instead, a method was proposed to synthesize PMMA from existing PMMA shapes as shown in Fig. 4.

Algorithm 2 Image translation

Input: V, P ▷ VerSe, PMMA set
Output: x_T such that $(y_T, x_T) \in T$ ▷ image of VerSe DT

1: $(y_v, x_v) = \texttt{RandomSelection}(V)$
2: $(y_p, x_p) = \texttt{RandomSelection}(P)$
3: $A_{pmma} = \texttt{ExtractPMMALabel}(x_p)$
4: $I_p = A_{pmma} \odot x_p$ ▷ extract PMMA image
5: $I'_p = \texttt{ResizeVolume}(I_p, x_v)$
6: $x_T = x_v + I'_p$

Algorithm 2 showed the pseudocode of the proposed image translation. In lines 1–2 of Algorithm 2, we sampled data pairs randomly from V and P, which referred to VerSe and PMMA set in our experiment. Next, A_{pmma}, the semantic label from the PMMA area of x_p, was extracted (line 3). Then I_p, the only voxels of PMMA area of x_p, was extracted by element-wise multiplication with the mask A_{pmma} (line 4). Before combining vertebral body image x_v and PMMA image I_p, we resized the volume of I_p to the size of x_v (line 5). Formally, we resized $I_p \in \mathbb{R}^{H' \times W' \times D'}$ to the size of $x_v \in \mathbb{R}^{H \times W \times D}$ so that $I'_p \in \mathbb{R}^{H \times W \times D}$. This shape matching step was devised to make PMMA be located in a realistic position of CT image x_v. Finally, we can create a new virtual PMMA inserted vertebral body image by combining vertebral body image of x_v and PMMA image I'_p (line 6). In other words, our virtual image mimicked the case in which the vertebral body of x_v has the same PMMA insertion surgery result of x_p.

By our image translation method, we can generate $n \times m$ virtual image pairs from n real data pairs from V and m real data pairs from P. Overall, a large source dataset V was domain transformed by label/image translation so that a large target set was made. In our experiment, set T, the domain transformed V, referred to VerSe DT.

C. Loss Function

We trained our model with two loss functions: dice loss $\mathcal{L}_d$ and shape-aware loss $\mathcal{L}_s$. Dice loss is defined as:

$$\mathcal{L}_d = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \varepsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \varepsilon}, \quad (2)$$

where the sums run over the N voxels, y is a ground truth label, $\hat{y}$ is a predicted value, and ε is a very small constant to prevent divided by zero (we use $\varepsilon = 1$). Shape-aware loss is modified from the one proposed in [20].

$$\mathcal{L}_s = -\frac{1}{N}\sum_{i=1}^{N} E_i \{ y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \}, \quad (3)$$

$$E_i = D(i_{pos}, B_{GT}). \quad (4)$$

Shape aware loss is the binary cross entropy loss with voxelwise weight E_i; E_i was computed by measuring the Euclidean distance between voxel i and the nearest boundary voxel of ground truth mask B_{GT}. Shape-aware loss [20] calculated Euclidean distance between prediction boundary and ground truth boundary, thus, Euclidean distance must be calculated in every epoch. However, the distance calculation of every 3D voxel in every epoch results in heavy computational cost. So, we modified it to be solely dependent on ground truth so that it can be pre-calculated only once before training. The shape loss term provided shape information of training data to the model. Finally, the total loss was described as the sum of dice loss and shape-aware loss:

$$\mathcal{L}_{total} = \mathcal{L}_d + \mathcal{L}_s. \quad (5)$$

IV. EXPERIMENTS

A. Dataset and Preprocessing

In this experiment, we used VerSe dataset as a large source dataset V, and PMMA set as a small target dataset P. The datasets were uniquely suited for our task as those dataset pairs have the domain shift in both image and label; VerSe contained vertebra label and vertebra image while PMMA set contains vertebral body label and PMMA inserted vertebral body (refer to Fig. 2). A large source dataset was changed by the domain. T was given the name VerSe DT (Domain Transformed VerSe). The composition for each dataset is given in Table 1. In this experiment, the small target dataset size was only 5% of the large source dataset. While training, every training

Table 1. Dataset split

	Train	Test
PMMA set	104	51
VerSe	2,334	1,107
VerSe DT	2,334	-

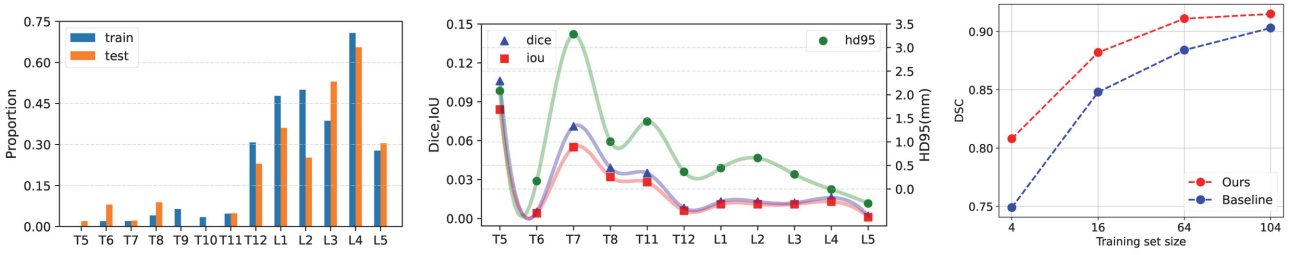

Fig. 5. (a) Data distribution of PMMA set. (b) Performance improvement of ours compared to the baseline. Left axis indicates dice score and IoU, Right axis indicates HD95. (c) Dice scores of the trained model depending on the target set size.

data was split into train and valid sets in the ratio of 8:2.

For a small target dataset, PMMA set, image data were obtained from those who were operated PMMA-injection surgery, i.e., percutaneous vertebroplasty, and were acquired at a local hospital (IRB approved dataset). Label annotation was done manually under the guidance of doctors. Data distribution of PMMA set in spine numbers, i.e., thoracic (T) to lumbar (L), is shown in Fig. 5(a). As we split the dataset into train and test randomly, the distribution of train and test is similar.

We used VerSe provided for MICCAI 2020 Large Scale Vertebrae Segmentation Challenge (VerSe 2020) [5]. Although each data in the original VerSe contained a whole spine CT image of a subject and segmentation labels in pairs, the data were reduced to contain only one vertebra each. And only thoracic and lumbar vertebra were reduced and used. This cropping makes the dataset have a reasonable domain shift from the target dataset, PMMA set.

For VerSe DT, theoretically we can generate 2,334 × 104 = 242,736 samples from 2,334 VerSe training set and 104 PMMA training set. Considering the computing time issue, we generated 2,334 samples for the training set, which included domain transformed data for each VerSe train data.

B. Implementation Details

Our segmentation deep learning network, 3D U-Net, consisted of contracting path and expansive path, each with five resolution steps. Dropout was included in the training step. Other details of network architecture were the same with 3D U-Net [3]. We used Adam optimizer with a learning rate of 0.001. We trained the model for 30,000 steps with a batch size of 2 on a single NVIDIA GeForce RTX 3090 GPU. Each optimization run was early stopped based on a validation error.

C. Evaluation Metrics

For evaluating the performance, we use three evaluation metrics. Given the ground truth G and the predicted segmentation P, we used the evaluation metrics: dice similarity coefficient (DSC), Intersection over Union (IoU), and 95% Hausdorff distance ($HD_{avg/95\%}$), defined by:

$$\mathrm{DSC} = 2\frac{|G \cap P| + \varepsilon}{|G| \cup |P| + \varepsilon},$$

$$\mathrm{IoU} = \frac{|G \cap P|}{|G \cup P|},$$

$$\mathrm{HD}_{avg/95\%} = \frac{\vec{d}_{H,95}(G,\ P) + \vec{d}_{H,95}(P,\ G)}{2},$$

$$\vec{d}_{H,95}(X,\ Y) = K_{95}\left(\min_{y \in Y}\ d(x,\ y)\right) \forall x \in X,$$

where $\varepsilon = 1$ and $d(x, y)$ denotes the distance between x and y, K_{95} denotes 95th percentile. We abbreviate $\mathrm{HD}_{avg/95\%}$ to $HD95$ in experimental results below.

D. Baseline and Model Variants

As a baseline model for our evaluation, we trained a network solely on the PMMA set. We referred to this model as baseline and used it as the starting point for all other variants. We compared the following approaches. The test score of the model in this paper was evaluated on the test set of PMMA set, to see a performance on our ultimate goal, PMMA-inserted vertebral body segmentation.

Merge with VerSe: A common approach to making use of a large dataset was simply to merge the dataset. In this approach, we merged PMMA set with VerSe and then trained the model from scratch.

Merge with VerSe DT: In this approach, we merged PMMA set and VerSe DT, domain transformed VerSe. This approach addressed the domain shift of VerSe regarding PMMA set. When merging with VerSe or VerSe DT, we oversampled PMMA set to deal with the data imbalance.

Transfer learning with VerSe: Another approach that utilized a large open dataset with domain shift is transfer learning. We compared two different approaches to transfer learning. First, in this approach, we pretrained

Table 2. Segmentation results on train from scratch with PMMA set (Merge)

Model	T5	T6	T7	T8	T11	T12	L1	L2	L3	L4	L5	All
Baseline	0.821	0.891	0.827	0.871	0.877	0.908	0.921	0.910	0.912	0.912	0.870	0.903
VerSe	0.909	0.877	0.881	0.894	0.899	0.904	0.928	0.910	0.910	0.925	0.878	0.909
VerSe DT (Ours)	0.904	0.897	0.881	0.860	0.905	0.917	0.933	0.920	0.923	0.926	0.892	0.915

Table 3. Segmentation results on transfer learning (Transfer)

Model	T5	T6	T7	T8	T11	T12	L1	L2	L3	L4	L5	All
Baseline	0.821	0.891	0.827	0.871	0.877	0.908	0.921	0.910	0.912	0.912	0.870	0.903
VerSe	0.914	0.897	0.870	0.902	0.905	0.910	0.931	0.924	0.922	0.928	0.875	0.916
VerSe DT (Ours)	0.927	0.896	0.898	0.910	0.912	0.916	0.934	0.923	0.924	0.928	0.872	0.920

the model with VerSe, and then fine-tuned the model with our target, PMMA set.

Transfer learning with VerSe DT: Second, in this approach, we pretrained the model with VerSe DT and then fine-tuned it with PMMA set. The two approaches only differed in the initialization of the fine-tuned model.

E. Results

We first compared the methods with or without our domain transformation in dice similarity score, as shown in Tables 2 and 3. We evaluated the performance for classes grouped according to spine numbers. The last column showed the overall dice score of each method.

Table 2 shows the results of baseline, Merge with VerSe, and Merge with VerSe DT. Our method with domain transformed VerSe, i.e., VerSe DT, showed clear improvement from the baseline and merge with VerSe on 9 out of 11 spine numbers.

Table 3 displays the results for transfer learning, which entails pre-training with VerSe or VerSe DT before fine-tuning with our target, the PMMA set. Transfer learning with VerSe DT shows the best result in overall spine numbers and outperforms in the class agnostic dice score. Clear and consistent improvement in each experimental setting with our method showed the effectiveness of our domain transformation.

1) Ablation study

In ablation study, we verified the effectiveness of our domain transformation methods and each step, i.e., label translation and image translation (Table 4). For transfer learning and merge, the performance keeps dropping without image translation and label translation. Further, we evaluated the training performance when there is no small target dataset *P*, i.e., PMMA set. The model trained with VerSe only showed 0.605 dice score on the test set of PMMA set, while our model trained with VerSe DT shows 0.862 dice score with high improvement. Likewise, the above ablation study showed, without PMMA set, the performance kept decreasing without image translation and label translation.

Table 4. Results of ablation study

Method	V	L	I	DSC	IoU	HD_{95} (mm)
Baseline				0.903	0.912	2.618
Transfer	√	√	√	0.920	0.926	2.099
	√			0.916	0.923	2.178
Merge	√	√	√	0.915	0.921	2.320
	√	√		0.912	0.919	2.358
	√			0.909	0.917	2.435
w/o PMMA set	√	√	√	0.862	0.880	3.416
	√	√		0.690	0.768	5.536
	√			0.605	0.728	7.016

For DSC and IoU, larger values mean closer to ground truth, and for HD_{95}, smaller values mean closer to ground truth.
V: VerSe set, L: label translation, I: image translation.

In addition, we demonstrated the effectiveness of our domain transformation according to target training set size (Fig. 5(c)). Merging with VerSe DT (Ours) consistently outperformed the baseline regardless of PMMA set size, and we further confirmed that the smaller the target PMMA set size is, the more gap between ours and the baseline exists.

2) Analysis on data distribution

We took a closer look at how class frequency affected the performance of our approach. Fig. 5(a) shows the train/test data distribution of PMMA set in proportion and Fig. 5(b) shows the performance improvement of our approach (Transfer with VerSe DT) over the baseline for each class. Spine numbers were plotted on the x-axis. The performance on three metrics was improved for most categories. Especially, our model was most effective for

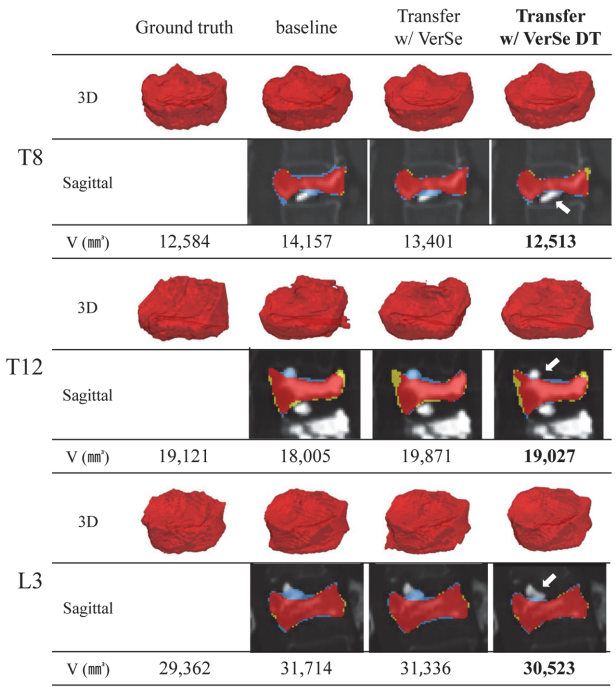

Fig. 6. Qualitative result. In sagittal view images, red indicates correct prediction of a model, blue indicates overestimation, and yellow indicates underestimation. V is a volume in the unit of mm^3.

the spine numbers of the least train set size (T5, T7). These results were consistent with the ablation study on the dataset size (Fig. 5(c)), that the effect of our domain transformation was greater as the target dataset size was smaller.

3) Qualitative evaluation

As shown in Fig. 6, we compared our method (Transfer learning with VerSe DT) with the baseline and transfer learning with VerSe. Compared with the two methods, our method inferred the boundaries of PMMA-inserted vertebral body boundary more accurately. To be specific, even in PMMA leakage—marked with white arrows—our method can discriminate vertebral body area and PMMA outside of the vertebral body. In addition, we investigated the shape and volume of each prediction, since the medical purpose of vertebral body segmentation was to measure accurate volume and shape. Compared with the two methods, we can estimate more reasonable 3D shapes of the vertebral bodies and infer the volume closer to the ground truth, which satisfies the medical purpose.

V. CONCLUSION

In this paper, we proposed a novel domain transformation for PMMA-inserted vertebral body segmentation. Our domain transformation framework included label translation and image translation, so that we can exploit a large source dataset to the target task, even when domain shift exists in both image and label. Experiments demonstrated that our domain transformation led to high-performance improvement through overall spine numbers. Further, various experimental results showed that our method was more effective when the target dataset size is smaller. Lastly, our domain transformation helped the segmentation model to infer a more accurate 3D shape and volume of PMMA-inserted vertebral body, which fulfilled the medical purpose of the segmentation.

CONFLICT-OF-INTEREST STATEMENT

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the Catholic University of Korea Daejeon St. Mary's Hospital (No. G01210193) and the National Research Foundation of Korea (No. NRF-2018X1A3A1069693).

The IRB document number is DIRB-20201123-003.

REFERENCES

1. A. Hatamizadeh, D. Yang, H. Roth, and D. Xu, "Unetr: transformers for 3D medical image segmentation," 2021 [Online]. Available: https://arxiv.org/abs/2103.10504.
2. C. Payer, D. Stern, H. Bischof, and M. Urschler, "Coarse to fine vertebrae localization and segmentation with SpatialConfiguration-Net and U-Net," in *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, Valletta, Malta, 2020, pp. 124-133.
3. O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham, Switzerland: Springer, 2016, pp. 424-432.
4. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993-2024, 2014.
5. A. Sekuboyina, M. E. Husseini, A. Bayat, M. Loffler, H. Liebl, H. Li, et al., "VerSe: a vertebrae labelling and segmentation benchmark for multi-detector CT images," *Medical Image Analysis*, vol. 73, article no. 102166, 2021. https://doi.org/10.1016/j.media.2021.102166
6. A. Komemushi, N. Tanigawa, S. Kariya, H. Kojima, Y. Shomura, and S. Sawada, "Percutaneous vertebroplasty for compression fracture: analysis of vertebral body volume by CT volumetry," *Acta Radiologica*, vol. 46, no. 3, pp. 276-279, 2005.

7. H. J. Lee, J. Park, I. W. Lee, J. S. Yi, and T. Kim, "Clinical, radiographic, and morphometric risk factors for adjacent and remote vertebral compression fractures over a minimum follow-up of 4 years after percutaneous vertebroplasty for osteoporotic vertebral compression fractures: novel three-dimensional voxel-based morphometric analysis," *World Neurosurgery*, vol. 125, pp. e146-e157, 2019. https://doi.org/10.1016/j.wneu.2019.01.020
8. W. Li, "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks," *Journal of Computer and Communications*, vol. 3, no. 11, article no. 61314, 2015. https://doi.org/10.4236/jcc.2015.311023
9. Q. Li, Z. Yu, Y. Wang, and H. Zheng, "TumorGAN: a multi-modal data augmentation framework for brain tumor segmentation," *Sensors*, vol. 20, no. 15, article no. 4203, 2020. https://doi.org/10.3390/s20154203
10. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer, 2015, pp. 234-241.
11. S. A. Harmon, T. H. Sanford, S. Xu, E. B. Turkbey, H. Roth, Z. Xu, et al., "Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets," *Nature Communications*, vol. 11, article no. 4080, 2020. https://doi.org/10.1038/s41467-020-17971-2
12. W. Xie, D. Jin, H. Ma, J. Ding, J. Xu, S. Zhang, and D. Liang, "Cement leakage in percutaneous vertebral augmentation for osteoporotic vertebral compression fractures: analysis of risk factors," *Clinical Spine Surgery*, vol. 29, no. 4, pp. E171-E176, 2016.
13. M. S. Aslan, A. Ali, D. Chen, B. Arnold, A. A. Farag, and P. Xiang, "3D vertebrae segmentation using graph cuts with shape prior constraints," in *Proceedings of 2010 IEEE International Conference on Image Processing*, Hong Kong, China, 2010, pp. 2193-2196.
14. G. Hille, S. Saalfeld, S. Serowy, and K. Tonnies, "Vertebral body segmentation in wide range clinical routine spine MRI data," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 93-99, 2018.
15. C. Chu, D. L. Belavy, G. Armbrecht, M. Bansmann, D. Felsenberg, and G. Zheng, "Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method," *PloS One*, vol. 10, no. 11, article no. e0143327, 2015. https://doi.org/10.1371/journal.pone.0143327
16. W. Yan, A. Sharma, and R. T. Tan, "Optical flow in dense foggy scenes using semi-supervised learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 2020, pp. 13256-13265.
17. S. Kim, Y. Huo, and S. E. Yoon, "Single image reflection removal with physically-based training images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 2020, pp. 5163-5172.
18. V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Scientific Reports*, vol. 9, article no. 16884, 2019. https://doi.org/10.1038/s41598-019-52737-x
19. A. Veit, N. Alldrin, G. Chechik, I. Krasin, A. Gupta, and S. Belongie, "Learning from noisy large-scale datasets with minimal supervision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 6575-6583.
20. S. M. Al Arif, K. Knapp, and G. Slabaugh, "Shape-aware deep convolutional neural network for vertebrae segmentation," in *Computational Methods and Clinical Applications in Musculoskeletal Imaging*. Cham, Switzerland: Springer, 2017, pp. 12-24.

Minyoung Park https://orcid.org/0000-0001-9534-4321

Minyoung Park received Master's degree in Software graduate program from the Korea Advanced Institute of Science and Technology (KAIST) in 2022. She received her Bachelor's degree in Biomedical Engineering and Computer Science and Engineering from Korea University in 2020. Her research interests include computer vision and medical image analysis.

Jinah Park https://orcid.org/0000-0003-4676-9862

Jinah Park received Ph.D. in Computer and Information Science from University of Pennsylvania in 1996. She is currently a Professor and KAIST ICT Endowed Chair Professor in the School of Computing, Korea Advanced Institute of Science and Technology (KAIST), leading the Computer Graphics and Visualization Research Lab since 2002. Her research interests include medical image data analysis, visualization, virtual reality and computer haptics.