Unsupervised Domain Adaptation for Pelvic Landmark Localization in CT and MR Using Landmark-Conditioned Synthesis*

Kangqing Ye, Simeng Luo, Xiaoru Gao, and Guoyan Zheng (✉)

Institute of Medical Robotics, School of Biomedical Engineering, Shanghai Jiao Tong University, No. 800, Dongchuan Road, Shanghai 200240, China
Corresponding Author: Guoyan Zheng (Email: guoyan.zheng@sjtu.edu.cn)

Abstract. Accurate 3D pelvic landmark localization is essential for hip morphology assessment and surgical planning. However, most models are trained on a single modality and therefore suffer substantial performance degradation when applied to a different modality due to severe domain shift. To alleviate this performance drop, unsupervised domain adaptation (UDA) methods based on image synthesis have been explored. Nevertheless, sparse landmark annotations make it difficult for existing synthesis methods to enforce semantic consistency between translated and real images, which limits the effectiveness of UDA. To address this limitation, we propose a UDA framework for 3D pelvic landmark localization using semantically consistent landmark-conditioned synthesis. Specifically, building on a CycleGAN backbone, we introduce landmark-conditioned adversarial learning (LCAL), which incorporates class- and laterality-conditioned local discriminators to promote semantically consistent image synthesis in regions surrounding each landmark. For localization, we employ modality-specific detectors to encourage consistent localization predictions across synthesized images of the same subject with different styles. Additionally, we propose a pseudo-label ensembling mechanism (PLEM) that produces reliable pseudo-labels by jointly considering localization confidence and cross-prediction spatial consistency, providing effective supervision in the target domain. Experiments on an unpaired hip CT-MR dataset demonstrate that our method outperforms other state-of-the-art UDA approaches on both CT-to-MR and MR-to-CT adaptation tasks. The code is available at https://github.com/KangqingYe/LCS-UDA.

Keywords: Landmark localization · Unsupervised domain adaptation · Image synthesis.

* This study was partially supported by the National Key R&D Program of China via project 2023YFB4706302 and by the National Natural Science Foundation of China via projects 62471293 and U20A20199.

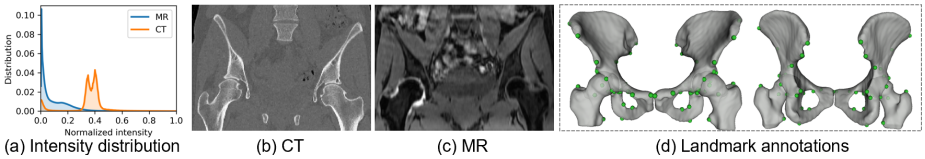

Fig. 1. Illustration of the CT-MR domain shift: intensity histograms (a), example CT and MR images (b-c), and landmark annotations (d).

1 Introduction

Accurate landmark localization in 3D pelvic images is a fundamental prerequisite for hip joint morphology assessment, thereby supporting diagnosis and preoperative planning for developmental hip dysplasia [16, 10]. With the advancement of deep learning, many automatic methods have been developed for 3D landmark localization, including regressing Gaussian heatmaps for landmark prediction [17, 5, 26], using a coarse-to-fine strategy to improve performance [27, 25], and incorporating auxiliary tasks to enforce geometric priors [8, 2].

Despite recent progress, most existing methods are designed for a single-domain setting, and their performance often degrades when applied to a different imaging modality (e.g., CT→MR). The degradation is mainly caused by domain shift, including differences in appearance and imaging characteristics [20]. Fig. 1 illustrates the intensity distributions and visual appearances of pelvic CT and MR images, along with a 3D visualization of landmark annotations. In practice, CT images are typically used to quantify acetabular morphology because of their excellent bone depiction quality [4], whereas MR images are used to assess soft-tissue pathology due to their superior soft-tissue contrast [12]. Consequently, annotating landmarks on osseous structures is generally more difficult and costly in MR images than in CT images [14]. To reduce the reliance on manual annotations, unsupervised domain adaptation (UDA) [1] for localization has been proposed, aiming to transfer knowledge from labeled source data to unlabeled target data.

Existing UDA approaches mainly involve three strategies: feature alignment, image synthesis, and pseudo-label-based learning [7]. Jin et al. [6] combined confidence-based pseudo-label selection and adversarial learning to facilitate feature alignment, and proposed a UDA localization method for X-ray images from different devices. However, the method's performance depends strongly on pseudo-label quality in the target domain. Large domain discrepancies between CT and MR may result in noisy pseudo-labels, causing significant performance degradation. In contrast, image synthesis methods [23, 21] generate target-style images from labeled source data and train localization models using corresponding source labels. For example, Zeng et al. [23] used CycleGAN for CT-MR synthesis and enforced segmentation consistency across real, synthesized, and reconstructed images to preserve semantic structures during translation. Nevertheless, preserving semantic consistency during image synthesis remains chal-

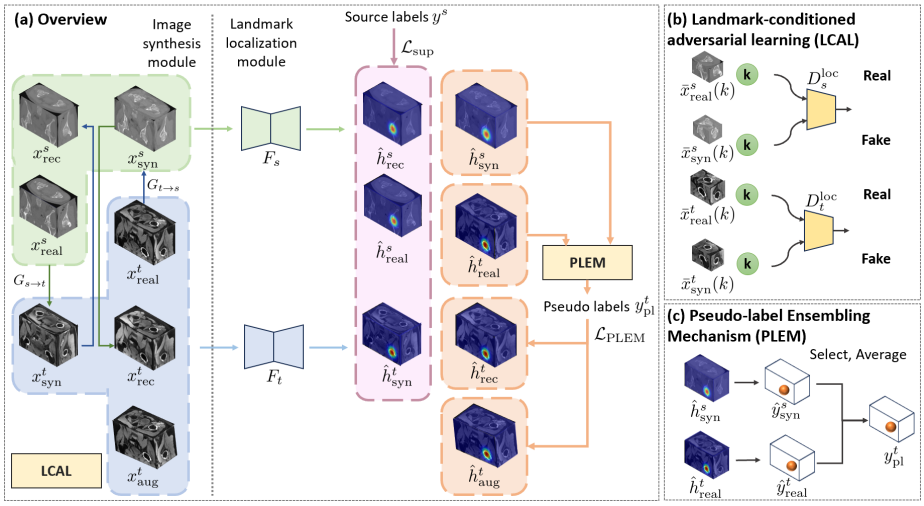

Fig. 2. A schematic illustration of our method. It consists of two modules: an image synthesis module with landmark-conditioned adversarial learning (LCAL) and a landmark localization module with the pseudo-label ensembling mechanism (PLEM).

lenging, especially with sparse landmark annotations. Diffusion-based synthesis has also been explored, but it is largely restricted to 2D slices due to computational cost [9, 24], and the few 3D extensions require paired data [15] or segmentations [3], neither of which is applicable to our setting. Moreover, these methods typically do not refine pseudo-labels in the target domain.

To address these limitations, we propose a UDA framework that integrates an image synthesis module with a landmark localization module. The image synthesis module builds on CycleGAN and introduces landmark-conditioned adversarial learning (LCAL), which applies class- and laterality-conditioned local discriminators to regions around each landmark to encourage semantically consistent synthesis. The landmark localization module uses two modality-specific detectors to enforce localization consistency across real, synthesized, and reconstructed images. We further propose a pseudo-label ensembling mechanism (PLEM) that refines target pseudo-labels using both localization confidence and spatial consistency across predictions, providing reliable supervision for the target domain. We validate our method on an unpaired hip CT-MR dataset. The results demonstrate the effectiveness of our method on both CT→MR and MR→CT adaptation tasks.

2 Method

In UDA landmark localization, we are given labeled source-domain images x^s with corresponding labels y^s and unlabeled target-domain images x^t. This work

aims to train a localization model that can achieve satisfactory performance on the target domain without access to target labels.

Our method consists of an image synthesis module and a landmark localization module, as shown in Fig. 2. In the image synthesis module, we adopt a standard CycleGAN backbone and introduce LCAL to facilitate semantically consistent synthesis. The landmark localization module includes two modality-specific landmark detectors (one per domain). These detectors are trained using source labels $y^s \in \mathbb{R}^{K\times 3}$ for images generated from the source images, and using pseudo-labels $y^t_{pl} \in \mathbb{R}^{K\times 3}$ for images generated from the target images, where K is the number of landmarks. The pseudo-labels y^t_{pl} and the corresponding validity mask $M_{pl} \in \{0,1\}^K$ are produced by PLEM.

2.1 Image Synthesis Module

The synthesis module is based on CycleGAN, with two generators ($G_{s\to t}$, $G_{t\to s}$) and two global discriminators (D^{glo}_s, D^{glo}_t). Given a real source image x^s_{real}, we synthesize $x^t_{\text{syn}} = G_{s\to t}(x^s_{\text{real}})$, and reconstruct $x^s_{\text{rec}} = G_{t\to s}(x^t_{\text{syn}})$ to enforce cycle consistency. Similarly, for a real target image x^t_{real}, we obtain $x^s_{\text{syn}} = G_{t\to s}(x^t_{\text{real}})$ and $x^t_{\text{rec}} = G_{s\to t}(x^s_{\text{syn}})$. For clarity, we define $\mathcal{L}^{\text{glo}}_G$ as the generator loss, $\mathcal{L}^{\text{glo}}_{D_s}$ and $\mathcal{L}^{\text{glo}}_{D_t}$ as the global discriminator losses for the source and target domains, respectively, and $\mathcal{L}_{\text{cyc}}$ as the cycle-consistency loss. In addition to the global adversarial learning in CycleGAN, we introduce LCAL to focus on improving generation quality in local regions surrounding landmarks.

Landmark-conditioned Adversarial Learning We introduce two local discriminators D^{loc}_s and D^{loc}_t for the source and target domains, respectively. To encourage semantically consistent synthesis around landmarks, we condition the local discriminator on the landmark class $k \in \{1, 2, ..., K\}$, making the adversarial supervision landmark-specific.

Since such supervision should be driven by the local context around each landmark, we apply each discriminator to a local patch cropped around the k-th landmark. For D^{loc}_s, we crop patches from x^s_{real} and x^s_{syn}, denoted by $\bar{x}^s_{\text{real}}(k)$ and $\bar{x}^s_{\text{syn}}(k)$ for the k-th landmark, respectively. However, unlabeled target images contain pseudo-labels for only a subset of landmarks. Using these incomplete pseudo-labels would result in training imbalance across iterations, which destabilizes training. To define crop centers for all landmarks, we assign proxy labels to landmarks without pseudo-labels using an anatomical prior [22, 19]. Although images are unpaired, all volumes are rigidly aligned to a common coordinate system. Therefore, anatomical structures are approximately consistent across subjects, and landmark coordinates are broadly comparable. We define the target crop center $\bar{y}^t_k$ as:

$$\bar{y}^t_k = \begin{cases} y^{t,k}_{\text{pl}}, & \text{if } M^k_{\text{pl}} = 1, \\ y^s_k, & \text{otherwise.} \end{cases} \tag{1}$$

Let $\mathcal{C}(x, y)$ denote a cubic cropping operator that extracts a local patch of size s_c centered at $y + r_c$, where r_c is a random offset. The inputs to D_s^{loc} are $\bar{x}^s_{\text{real}}(k) = \mathcal{C}(x^s_{\text{real}}, y^s_k)$ and $\bar{x}^s_{\text{syn}}(k) = \mathcal{C}(x^s_{\text{syn}}, \bar{y}^t_k)$. Similarly, the inputs to D_t^{loc} are $\bar{x}^t_{\text{real}}(k) = \mathcal{C}(x^t_{\text{real}}, \bar{y}^t_k)$ and $\bar{x}^t_{\text{syn}}(k) = \mathcal{C}(x^t_{\text{syn}}, y^s_k)$.

To encode the landmark class while accounting for left-right symmetry, we decompose each landmark class k into a laterality-agnostic landmark class $\bar{k} \in \{1, 2, ..., \frac{K}{2}\}$ and a laterality indicator $b \in \{-1, 1\}$ (e.g., $b = +1$ for left and $b = -1$ for right). The landmark-class embedding $\mathbf{e}(k)$ is defined as:

$$\mathbf{e}(k) = \bar{\mathbf{e}}(\bar{k}) + b \cdot \mathbf{v}, \tag{2}$$

where $\bar{\mathbf{e}}(\cdot)$ is a learnable embedding table and $\mathbf{v} \in \mathbb{R}^d$ is a learnable vector of dimension d representing laterality information.

Following the projection discriminator [13], the local discriminator encodes the input image $\bar{x}$ using a small convolutional neural network with global pooling to obtain a feature vector $\mathbf{d} \in \mathbb{R}^d$, and outputs the conditional score:

$$D^{\text{loc}}(\bar{x}, k) = \sigma(\mathbf{w}^\top \mathbf{d} + \mathbf{e}(k)^\top \mathbf{d}), \tag{3}$$

where $\mathbf{w} \in \mathbb{R}^d$ is a learnable weight vector and σ denotes the sigmoid function.

We train D_s^{loc} and D_t^{loc} using adversarial learning. For each domain $m \in \{s, t\}$ and each landmark k, the discriminator maximizes $\mathcal{L}^{\text{LCAL}}_{D_m}$ and the generators minimize $\mathcal{L}^{\text{LCAL}}_G$:

$$\begin{aligned}
\mathcal{L}^{\text{LCAL}}_{D_m} &= \sum_{k=1}^{K} \left(\log D^{\text{loc}}_m\left(\bar{x}^m_{\text{real}}(k), k\right) + \log\left(1 - D^{\text{loc}}_m\left(\bar{x}^m_{\text{syn}}(k), k\right)\right)\right), \\
\mathcal{L}^{\text{LCAL}}_G &= \sum_{m \in \{s,t\}} \sum_{k=1}^{K} \left[- \log D^{\text{loc}}_m\left(\bar{x}^m_{\text{syn}}(k), k\right)\right].
\end{aligned} \tag{4}$$

2.2 Landmark Localization Module

We employ two landmark detectors, one for the source domain (F_s) and one for the target domain (F_t). Each detector is supervised using Gaussian heatmaps h. Predicted coordinates $\hat{y} = \{\hat{y}^k\}_{k=1}^K$ are obtained by taking the argmax of the predicted heatmaps $\hat{h} = \{\hat{h}^k\}_{k=1}^K$. To enforce consistent localization predictions across real images and their synthesized and reconstructed counterparts, we define the source-domain supervised loss as:

$$\mathcal{L}_{\text{sup}} = \frac{1}{K} \sum_{k=1}^{K} \Big(\left\| \hat{h}^{s,k}_{\text{real}} - h^{s,k} \right\|_2^2 + \left\| \hat{h}^{t,k}_{\text{syn}} - h^{s,k} \right\|_2^2 + \left\| \hat{h}^{s,k}_{\text{rec}} - h^{s,k} \right\|_2^2 \Big). \tag{5}$$

For target images, their generated counterparts are supervised using pseudo-labels produced by PLEM.

Pseudo-label Ensembling Mechanism To obtain reliable pseudo-labels for target images, we ensemble two predictions for each target image: one from the real target image, $\hat{y}^{t,k}_{\text{real}}$, and the other from the corresponding synthesized image $\hat{y}^{s,k}_{\text{syn}}$. As pseudo-labels may be noisy due to domain shift and synthesis errors, we retain pseudo-labels that are both confident and spatially consistent. Confidence is defined as the maximum response of each predicted heatmap.

We define a confidence mask M^k_c and a spatial consistency mask M^k_d, and then obtain the final validity mask $M^k_{\text{pl}} = M^k_c \cdot M^k_d$:

$$M^k_c = \mathbb{I}\Big[\max(\hat{h}^{t,k}_{\text{real}}) > \tau_c\Big] \cdot \mathbb{I}\Big[\max(\hat{h}^{s,k}_{\text{syn}}) > \tau_c\Big], M^k_d = \mathbb{I}\Big[\Big\|\hat{y}^{t,k}_{\text{real}} - \hat{y}^{s,k}_{\text{syn}}\Big\|_2 < \tau_d\Big], \tag{6}$$

where $\mathbb{I}[\cdot]$ is the indicator function. The pseudo-labels are obtained by averaging the two predictions: $y^{t,k}_{\text{pl}} = \frac{1}{2}\left(\hat{y}^{t,k}_{\text{real}} + \hat{y}^{s,k}_{\text{syn}}\right)$.

Following the weak-to-strong consistency learning [18], we generate an augmented target image x^t_{aug} via a spatial augmentation transform $\mathcal{T}$. Let $h^{t,k}_{\text{pl}}$ denote the heatmap generated from $y^{t,k}_{\text{pl}}$. With predictions $\hat{h}^{t,k}_{\text{rec}}$ on x^t_{rec} and $\hat{h}^{t,k}_{\text{aug}}$ on x^t_{aug}, we adopt a masked MSE loss so that only valid landmarks contribute:

$$\mathcal{L}_{\text{PLEM}} = \frac{1}{\sum_{k=1}^{K} M^k_{\text{pl}}} \sum_{k=1}^{K} M^k_{\text{pl}} \Big(\Big\|\hat{h}^{t,k}_{\text{rec}} - h^{t,k}_{\text{pl}}\Big\|^2_2 + \Big\|\hat{h}^{t,k}_{\text{aug}} - \mathcal{T}\Big(h^{t,k}_{\text{pl}}\Big)\Big\|^2_2 \Big). \tag{7}$$

2.3 Training strategy

We adopt a three-stage training schedule for stability: (1) train generators and discriminators using the standard CycleGAN objective and LCAL; (2) freeze the synthesis networks and train the landmark detectors using $\mathcal{L}_{\text{sup}}$ and $\mathcal{L}_{\text{PLEM}}$; (3) train the entire network. The discriminator maximizes $\mathcal{L}_D$ and the generators and landmark detectors minimize $\mathcal{L}_{\text{total}}$:

$$\begin{aligned} \mathcal{L}_D &= \mathcal{L}^{\text{glo}}_{D_s} + \mathcal{L}^{\text{glo}}_{D_t} + \mathcal{L}^{\text{LCAL}}_{D_s} + \mathcal{L}^{\text{LCAL}}_{D_t}, \\ \mathcal{L}_{\text{total}} &= \mathcal{L}^{\text{glo}}_{G} + \lambda_{\text{cyc}}\mathcal{L}_{\text{cyc}} + \lambda_{\text{LCAL}}\mathcal{L}^{\text{LCAL}}_{G} + \mathcal{L}_{\text{sup}} + \mathcal{L}_{\text{PLEM}}. \end{aligned} \tag{8}$$

3 Experiments

3.1 Dataset and Evaluation Metric

We constructed an unpaired hip CT-MR dataset after local institution review board (IRB) approval (Approval No. 16-07-13, the teaching and research management of Inselspital, University of Bern, Switzerland). The dataset contained 70 CT images and 68 MR images. Each image was annotated with 54 anatomical landmarks for morphology assessment [11]. We randomly selected 14 and 7 CT-MR pairs for testing and validation respectively, while the remaining 49 CT and 47 MR images were used for training (in an unpaired manner). All images

Table 1. Results of our methods and other approaches under UDA settings.

Method Type	Methods	CT→MR				MR→CT			
		MRE↓	SDR (%)↑			MRE↓	SDR (%)↑		
		(mm)	3mm	5mm	10mm	(mm)	3mm	5mm	10mm
Lower bound	No adaptation	94.75	1.5	5.8	15.2	24.18	13.0	33.1	76.5
Upper bound	Fully supervised	3.36	52.0	84.4	99.1	2.78	66.7	90.0	98.9
UDA	CycleGAN [28]	4.13	33.9	73.9	97.6	4.61	21.2	63.5	97.8
	ICMSC [23]	3.92	36.4	76.1	98.7	4.24	28.8	72.1	98.5
	UDA-landmark [6]	93.03	2.4	6.6	20.4	19.16	11.0	29.9	77.1
	UMSCS [21]	4.27	29.4	71.3	97.8	4.33	26.5	70.4	97.6
	Ours	**3.53**	**46.0**	**83.3**	**99.1**	**3.85**	**36.0**	**80.6**	**98.4**

were resampled to a uniform voxel spacing of $1.5 \times 1.5 \times 1.5$ mm^3 and cropped to a size of $250 \times 159 \times 160$ following paper [19].

We adopted two commonly used metrics [6, 8] as follows: (1) Mean Radial Error (MRE), the mean Euclidean distance between the predicted and ground-truth landmarks; and (2) Successful Detection Rate (SDR), the percentage of landmarks correctly detected within thresholds of 3.0 mm, 5.0 mm, and 10.0 mm from the ground-truth.

3.2 Implementation Details

We used SCN [17] as the landmark detector. Input patches of $80 \times 80 \times 80$ voxels were extracted. For LCAL, local patches of side length $s_c = 40$ were cropped with random offsets $r_c \in [-3, 3]^3$. We set $\lambda_{\text{cyc}} = 10$, $\lambda_{\text{LCAL}} = 0.5$, $d = 512$, $\tau_c = 0.9$, and $\tau_d = 3$ empirically. The spatial augmentation $\mathcal{T}$ included random rotation within $[-10°, 10°]$ and random scaling within [0.9, 1.1]. We trained the three stages for 200, 200, and 600 epochs, respectively, using the Adam optimizer with a cosine-annealed learning rate from 2×10^{-4} to 2×10^{-5}. All experiments were conducted on an NVIDIA RTX 4090 GPU.

3.3 Comparison with SOTA Approaches

We compared our method with the following state-of-the-art (SOTA) UDA approaches on two tasks: CT→MR and MR→CT. CycleGAN [28] translated source images to the target style, and then trained a landmark detector on the synthetic images. UDA-landmark [6], originally proposed for X-ray images, was adapted to our 3D setting. We also included two CycleGAN-based UDA segmentation methods, ICMSC [23] and UMSCS [21]. For fairness, all methods used the same landmark detector architecture. We additionally reported a source-only baseline (trained on the source domain and directly tested on the target domain) as a lower bound, and a model trained with target ground-truth landmarks as an upper bound.

The comparison results are shown in Table 1. Our method consistently outperformed other competing approaches on both tasks, achieving an MRE of 3.53

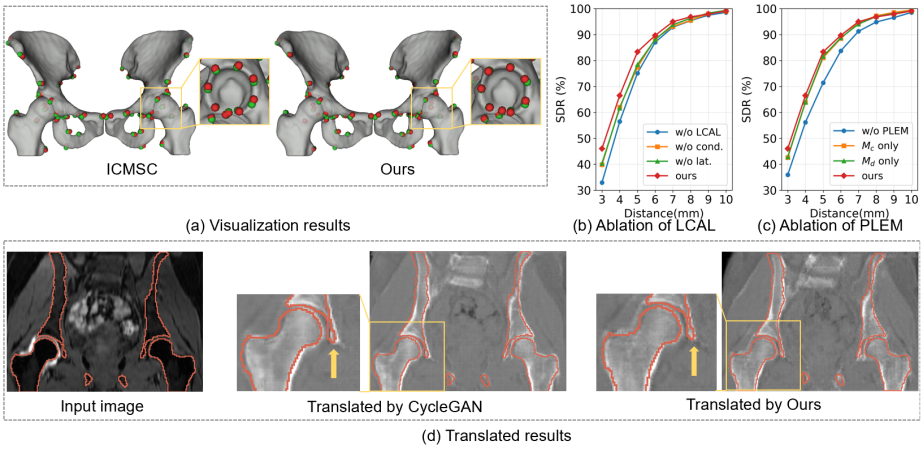

Fig. 3. (a) Visualization results from our method and the second-best method on the CT→MR task, with ground-truth landmarks (green) and predicted landmarks (red). (b) Ablation results of LCAL. "w/o cond." denotes local discriminators without class or laterality conditioning; "w/o lat." removes laterality conditioning while retaining class conditioning. (c) Ablation results of PLEM. (d) Images synthesized by CycleGAN and our method. The osseous segmentation of the input image is shown in red.

mm on the CT→MR task, and an MRE of 3.85 mm on the MR→CT task. The qualitative results of our method and the second-best method ICMSC on the task CT→MR are shown in Fig. 3(a).

3.4 Ablation Studies

We conducted ablation studies to evaluate the impact of LCAL and PLEM, as shown in Table 2. To assess the contribution of PLEM, we replaced the ensembled pseudo-labels y^t_{pl} with $\hat{y}^t_{\text{real}}$ while keeping other settings unchanged. Removing both LCAL and PLEM resulted in the lowest performance, indicating that both components are essential.

We further analyzed the components of LCAL and PLEM on the CT→MR task. For LCAL, class- and laterality-conditioned local discriminators achieved the best accuracy compared with variants without class conditioning or laterality conditioning, as shown in Fig. 3(b). For PLEM, we evaluated the contribution of masks M_c and M_d. The variant using both masks performed the best (Fig. 3(c)), suggesting complementary filtering based on confidence and spatial consistency. Finally, we visualized and compared the synthesized images produced by CycleGAN and our method. As shown in Fig. 3(d), CycleGAN altered anatomical structures around landmarks (e.g., the acetabular notch), whereas our method better preserved semantic consistency between translated and real images.

Table 2. Ablation results on the key components of our method.

LCAL	PLEM	CT→MR				MR→CT			
		MRE↓	SDR (%)↑			MRE↓	SDR (%)↑		
		(mm)	3mm	5mm	10mm	(mm)	3mm	5mm	10mm
		4.39	29.8	66.7	97.5	5.23	12.8	54.0	96.4
✓		4.04	36.0	71.4	98.5	4.66	23.4	62.1	97.5
	✓	4.00	33.1	75.1	98.5	4.45	19.7	70.0	98.3
✓	✓	**3.53**	**46.0**	**83.3**	**99.1**	**3.85**	**36.0**	**80.6**	**98.4**

4 Conclusion

In this paper, we presented a novel UDA framework for 3D landmark localization that integrated landmark-conditioned image synthesis with pseudo-label ensembling. LCAL promoted semantically consistent synthesis around landmarks, while PLEM provided reliable target-domain supervision, which together led to improved cross-modality localization. Experiments on an unpaired hip CT-MR dataset demonstrated the effectiveness of our approach, providing a practical solution for UDA landmark localization.

Disclosure of Interests. The authors declare no conflict of interest.

References

1. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189. PMLR (2015)
2. Gong, H., Wan, B., Kang, L., Wan, X., Zhang, L., Li, H.: Boundary as the bridge: Towards heterogeneous partially-labeled medical image segmentation and landmark detection. IEEE Transactions on Medical Imaging (2025)
3. Gong, H., Wang, Y., Wang, Y., Xiao, J., Wan, X., Li, H.: Diffuse-uda: Addressing unsupervised domain adaptation in medical image segmentation with appearance and structure aligned diffusion models. arXiv preprint arXiv:2408.05985 (2024)
4. Griffin, D.R., Dickenson, E.J., O'donnell, J., Awan, T., Beck, M., Clohisy, J., Dijkstra, H., Falvey, E., Gimpel, M., Hinman, R., et al.: The warwick agreement on femoroacetabular impingement syndrome (fai syndrome): an international consensus statement. British journal of sports medicine **50**(19), 1169–1176 (2016)
5. Huang, Z., Tang, T., Xu, R., Wei, Y., Yang, W., Wang, S., Sun, X., Li, H., Yao, Q.: H3de-net: Efficient and accurate 3d landmark detection in medical imaging. arXiv preprint arXiv:2502.14221 (2025)
6. Jin, H., Che, H., Chen, H.: Unsupervised domain adaptation for anatomical landmark detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 695–705. Springer (2023)
7. Kumari, S., Singh, P.: Deep learning for unsupervised domain adaptation in medical imaging: Recent advancements and future perspectives. Computers in Biology and Medicine **170**, 107912 (2024)
8. Li, X., Lv, S., Li, M., Zhang, J., Jiang, Y., Qin, Y., Luo, H., Yin, S.: Sdmt: spatial dependence multi-task transformer network for 3d knee mri segmentation and landmark localization. IEEE transactions on medical imaging **42**(8), 2274–2285 (2023)

9. Li, Y., Shao, H.C., Liang, X., Chen, L., Li, R., Jiang, S., Wang, J., Zhang, Y.: Zero-shot medical image translation via frequency-guided diffusion models. IEEE transactions on medical imaging **43**(3), 980–993 (2023)
10. Liu, L., Ecker, T., Schumann, S., Siebenrock, K., Nolte, L., Zheng, G.: Computer assisted planning and navigation of periacetabular osteotomy with range of motion optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 643–650. Springer (2014)
11. Liu, L., Siebenrock, K., Nolte, L.P., Zheng, G.: Computer-assisted planning, simulation, and navigation system for periacetabular osteotomy. In: Intelligent Orthopaedics: artificial intelligence and smart image-guided Technology for Orthopaedics, pp. 143–155. Springer (2018)
12. Mascarenhas, V.V., Rego, P., Dantas, P., Morais, F., McWilliams, J., Collado, D., Marques, H., Gaspar, A., Soldado, F., Consciencia, J.G.: Imaging prevalence of femoroacetabular impingement in symptomatic patients, athletes, and asymptomatic individuals: a systematic review. European journal of radiology **85**(1), 73–95 (2016)
13. Miyato, T., Koyama, M.: cgans with projection discriminator. arXiv preprint arXiv:1802.05637 (2018)
14. Morbée, L., Chen, M., Van Den Berghe, T., Schiettecatte, E., Gosselin, R., Herregods, N., Jans, L.B.: Mri-based synthetic ct of the hip: can it be an alternative to conventional ct in the evaluation of osseous morphology? European radiology **32**(5), 3112–3120 (2022)
15. Pan, S., Abouei, E., Wynne, J., Chang, C.W., Wang, T., Qiu, R.L., Li, Y., Peng, J., Roper, J., Patel, P., et al.: Synthetic ct generation from mri using 3d transformer-based denoising diffusion model. Medical Physics **51**(4), 2538–2548 (2024)
16. Park, J., Kim, J.Y., Kim, H.D., Kim, Y.C., Seo, A., Je, M., Mun, J.U., Kim, B., Park, I.H., Kim, S.Y.: Analysis of acetabular orientation and femoral anteversion using images of three-dimensional reconstructed bone models. International journal of computer assisted radiology and surgery **12**(5), 855–864 (2017)
17. Payer, C., Štern, D., Bischof, H., Urschler, M.: Integrating spatial configuration into heatmap regression based cnns for landmark localization. Medical image analysis **54**, 207–219 (2019)
18. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
19. Wang, R., Heimann, A.F., Tannast, M., Zheng, G.: Cyclesgan: A cycle-consistent and semantics-preserving generative adversarial network for unpaired mr-to-ct image synthesis. Computerized Medical Imaging and Graphics **117**, 102431 (2024)
20. Wilson, G., Cook, D.J.: A survey of unsupervised deep domain adaptation. ACM Transactions on Intelligent Systems and Technology (TIST) **11**(5), 1–46 (2020)
21. Yang, F., Li, X., Wang, B., Teng, P., Liu, G.: Umscs: a novel unpaired multimodal image segmentation method via cross-modality generative and semi-supervised learning. International Journal of Computer Vision **133**(7), 4442–4464 (2025)
22. Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Prince, J.L., Xu, Z.: Unsupervised mr-to-ct synthesis using structure-constrained cyclegan. IEEE transactions on medical imaging **39**(12), 4249–4261 (2020)
23. Zeng, G., Lerch, T.D., Schmaranzer, F., Zheng, G., Burger, J., Gerber, K., Tannast, M., Siebenrock, K., Gerber, N.: Semantic consistent unsupervised domain adaptation for cross-modality medical image segmentation. In: International Conference

on Medical Image Computing and Computer-Assisted Intervention. pp. 201–210. Springer (2021)
24. Zeng, H., Zou, K., Chen, Z., Gao, Y., Chen, H., Zhang, H., Zhou, K., Wang, M., Jiang, C., Goh, R.S.M., et al.: Training-free image style alignment for domain shift on handheld ultrasound devices. IEEE Transactions on Medical Imaging **44**(4), 1942–1952 (2024)
25. Zhai, H., Chen, Z., Li, L., Tao, H., Wang, J., Li, K., Shao, M., Cheng, X., Wang, J., Wu, X., et al.: Two-stage multi-task deep learning framework for simultaneous pelvic bone segmentation and landmark detection from ct images. International Journal of Computer Assisted Radiology and Surgery **19**(1), 97–108 (2024)
26. Zhao, X., Xiao, D., Zhang, T., Fan, J., Ai, D., Fu, T., Lin, Y., Shao, L., Song, H., Wang, J., et al.: Defect-adaptive landmark detection in pelvis ct images via personalized structure-aware learning. Computerized Medical Imaging and Graphics p. 102693 (2025)
27. Zhu, H., Li, X., He, T., Liu, B., Tang, W., Guo, J.: Enhanced landmark detection in oral and maxillofacial ct and cbct scans: A multi-stage global-local integration approach. IEEE Transactions on Instrumentation and Measurement (2025)
28. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)