Image-to-Graph Convolutional Network for 2D/3D Deformable Model Registration of Low-Contrast Organs

Megumi Nakao, *Member, IEEE*, Mitsuhiro Nakamura, and Tetsuya Matsuda, *Member, IEEE*

***Abstract*— Organ shape reconstruction based on a single-projection image during treatment has wide clinical scope, e.g., in image-guided radiotherapy and surgical guidance. We propose an image-to-graph convolutional network that achieves deformable registration of a three-dimensional (3D) organ mesh for a low-contrast two-dimensional (2D) projection image. This framework enables simultaneous training of two types of transformation: from the 2D projection image to a displacement map, and from the sampled per-vertex feature to a 3D displacement that satisfies the geometrical constraint of the mesh structure. Assuming application to radiation therapy, the 2D/3D deformable registration performance is verified for multiple abdominal organs that have not been targeted to date, i.e., the liver, stomach, duodenum, and kidney, and for pancreatic cancer. The experimental results show shape prediction considering relationships among multiple organs can be used to predict respiratory motion and deformation from digitally reconstructed radiographs with clinically acceptable accuracy.**

***Index Terms*— Deep learning, deformable registration, graph convolutional network, abdominal organs, low-contrast images.**

I. Introduction

ORGAN positions and shapes from three-dimensional (3D) medical images constitute patient-specific morphological information that is essential to diagnosis and pre-treatment planning. However, organs may move or deform during surgical treatment or through several weeks of radiation therapy [1], [2]. Post-imaging time-series shape changes in organs can prevent tumor localization and hinder treatment. Existing imaging devices for use during treatment have certain limitations; thus, two-dimensional (2D) images facilitating real-time measurements (e.g., endoscopic and X-ray images) are available but 3D imaging is limited [3]–[5].

Organ shape reconstruction based on a single-projection image during treatment has wide clinical scope including image-guided therapy/intervention. However, this problem is ill-posed without prior knowledge as it requires transformation of 2D space points into points in a higher-dimensional space. One solution is 2D/3D registration, which involves the patient-specific organ shape from dense 3D computed tomography (CT) or magnetic resonance imaging (MRI) images taken prior to treatment and use of these data as prior knowledge. This approach aims to solve alignment and deformation of the organ-shape models to 2D projection images in real time, and has undergone intensive research in the field of medical image analysis over the past decade [6], [7]. In particular, many studies have examined rigid-body 2D/3D image registration [8], [9] as an optimization problem for parameter sets that determine the position and orientation.

2D/3D deformable registration of soft organs requires local point-to-point correspondence between 2D images and 3D volumes. Unlike rigid-body registration, large-scale parameters must be optimized proportional to the number of sampling points. Deformable registration between 3D volumes poses a similar problem [10]. Diffeomorphic mapping-based regularization [11], [12] enables calculation of a displacement field that can obtain smooth correspondence between sampling points; however, pairwise optimization has high computational cost for a large-scale parameter set. Thus, recent studies have investigated 3D displacement field learning using a convolutional neural network (CNN) [13]–[18]. Notably, machine learning models trained via parallel computing with a graphics processing unit can provide accelerated registration.

Single-image-based 2D/3D deformable registration has less constraints than the above-mentioned registration between 3D volumes, making stable optimization difficult. Predictions based on input images alone have high uncertainty, and the mapping between the organ shape model and 2D images, along with its learning method, are key. Some studies use bi-planar X-ray images to improve prediction accuracy [19], [20]. In the

Manuscript received 11 June 2022; revised 21 July 2022; accepted 25 July 2022. Date of publication 28 July 2022; date of current version 2 December 2022. This work was supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) under Grant 22H03021 and Grant 19H04484. *(Corresponding author: Megumi Nakao.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board in Kyoto University under Application No. R1499, and performed in line with the Declaration of Helsinki.

Megumi Nakao and Tetsuya Matsuda are with the Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (e-mail: megumi@i.kyoto-u.ac.jp; tetsu@i.kyoto-u.ac.jp).

Mitsuhiro Nakamura is with the Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan (e-mail: m_nkmr@kuhp.kyoto-u.ac.jp).

Digital Object Identifier 10.1109/TMI.2022.3194517

field of computer vision, human posture and various general objects have been investigated, with a camera image database corresponding to a 3D shape being used as a background [21], [22]. Recent works have proposed integrating the CNN and a graph convolutional network (GCN) [23], or an estimation framework that is robust against occlusion through a self-attention network [24].

As regards medical imaging, collection of organ deformation data paired with 2D images is difficult, and few cases have been reported to date. A learning method for a registration map involving correspondence between an area on a 2D projection image and a local area in a 3D volume using a CNN has been proposed [25]–[27]. Additionally, for 2D/3D deformable registration of soft organs for surgical guidance, model-based optimization for endoscopic images has been attempted [28]–[32]. However, to the best of our knowledge, no studies have provided a framework for deep learning-based 2D/3D deformable model registration of abdominal soft organs. Within the scope of our survey, no empirical cases using actual patient data have been reported.

This study introduces an image-to-graph convolutional network that enables 3D organ-shape reconstruction and localization based on a low-contrast projection image. The proposed network provides a new end-to-end framework that achieves real-time 2D/3D deformable registration through integration of an image-based generative network and GCN. The generative network learns the transformation from the 2D projection image to a displacement map based on pairwise 3D meshes obtained before and after deformation. The GCN samples the input features of each node from the generated displacement field and learns the transformation into a final 3D displacement vector that satisfies the geometrical constraints. Finally, our network outputs a 3D mesh, the position and deformation of which are registered to the input 2D projection image.

Assuming application to radiation therapy, the shape reconstruction performance from a single 2D projection image targeting the abdominal organs of actual patients is verified experimentally. This is the first study to demonstrate 2D/3D deformable registration of the liver, duodenum, and kidney, and the gross tumor volume (GTV) of pancreatic cancer. Many variations in organ shape and deformation exist between patients, and there are almost no visual clues (such as contours) in the low-contrast 2D projection images (Fig. 1); thus, accurate prediction of the organ positions and shapes was previously considered difficult. We also show that respiratory dynamics and deformation can be predicted from digitally reconstructed radiograph (DRR) images via statistical data augmentation for 3D-CT and simultaneous prediction of multiple organ shapes.

The methods reported herein extend a previous model-based deformable registration network (IGCN) [33] presented at the 2021 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). In the image-to-graph convolutional networks, image features are obtained from the reference points by projecting discrete and spatially discontinuous vertices of a mesh. Pixel2Mesh [22], which was developed for general 3D objects, uses CNNs to extract image features for shape reconstruction. IGCN [33] improved registration accuracy by obtaining more effective features using warped reference points, where the relationship between the extracted features and the 3D deformation of the mesh was learned. However, registration errors remain because of the instability in learning image features, i.e., the optimization of the weights in the CNNs and GCNs. We focused on the difference between feature extraction in CNNs and sparse references in GCNs, and considered that spatially discrete references make it difficult to obtain the smooth gradients necessary to optimize the weights in CNNs. Our new framework (called IGCN+ in this paper) aims to solve this problem by orienting the image transformation network so that it can learn to generate the pixel-level, dense deformation map.

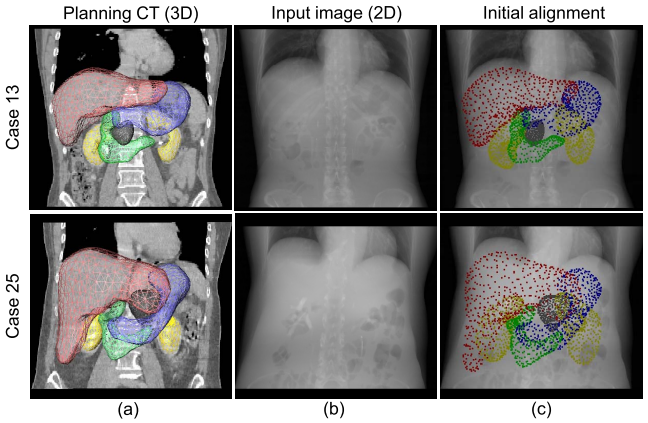

Fig. 1. Problem definition for 2D/3D deformable model registration for abdominal organs. (a) Pre-treatment 3D-CT and surface meshes of liver, stomach, duodenum, kidneys, and pancreatic GTV. 2D projection images (b) of target states and (c) overlaid with projected vertices of pre-treatment CT meshes.

Secondly, we propose the multiple-organ reconstruction framework and investigate its effectiveness for learning deformable shape registration. In the previous paper [33], we focused on the shape reconstruction of a single organ and reported liver reconstruction results as a preliminary study. This paper shows that the multiple-organ reconstruction framework learns the relationship between the position and shape of abdominal organs and improves registration accuracy. We newly validated the registration performance on the stomach, duodenum, kidney, and a GTV of pancreatic cancer, and confirmed that clinically acceptable accuracy could be achieved. We also investigated the role of the statistical generative models to augment respiratory deformation datasets.

In summary, the contributions of this study are as follows:

- A new 2D/3D deformable model registration framework that integrates a pixel-level deformation map generator and GCN;
- Simultaneous shape reconstruction of five abdominal organs, the contours of which are not directly visible on a 2D projection image;
- Application to localization of GTV and organ-at-risk (OAR) volumes assuming dynamic tumor-tracking radiotherapy with clinically acceptable estimation accuracy.

II. Related Work

A. Optimization-Based Approaches

Optimization-based 2D/3D registration has been extensively researched over the past 20 years [6], [34]–[37]. Rigid-body and deformable registration involve formulation as a transformation matrix or as a displacement vector field optimization problem, respectively [8], [38], [39]. Because of the high density and large scale of 3D volumes in particular, approaches that construct shape models based on anatomical labels and seek positions and deformation on 2D images have been very successful. Notably, mesh-based shape representation can express organ elastic properties [30], [40], [41] and statistical shape variations [1], [42], [43] with few variables and high computational efficiency. Deformable model registration of X-ray and endoscopic images has been attempted, with the aim of surgical guidance [29]–[32]. In radiation therapy, contour definitions and statistical atlases of GTVs and OARs [1], [7], [44], [45] can be directly applied in model-based registration. However, 2D/3D registration based on model parameter optimization is limited to the subspace of physical models and statistical shape models defined in advance by the expressible shape variations. Additionally, the high degrees of freedom of the position and deformation in a 2D image make it difficult to set objective functions from which a stable solution can be derived, and each registration requires time for iterative optimization calculations.

B. Learning-Based Approaches

Recently, deep learning-based frameworks have attracted attention [46]–[48] because of the high uncertainty and computational costs inherent in model-based 2D/3D deformable registration. Pointnet [21], a CNN-based framework that generates a 3D point cloud from a single-viewpoint image, was applied to 2D virtual images of statistical pneumothorax lung models, and lobe shapes were reconstructed [47]. However, in point cloud representation, surface and topological information on the inter-vertex relationships, which are important for deformation field computation, are lost. Wang at al. proposed Pixel2Mesh (P2M) [22] to generate a 3D mesh from a 2D projection image. P2M uses latent image features to deform an ellipsoid template into the target shape. A recent work [48] was the first to apply P2M to respiratory deformation estimation from a DRR, with 3D lung shapes being artificially generated from multiple initial 3D templates with free-form deformation. We previously implemented 2D/3D deformable registration methods using four-dimensional (4D) CT data for real patients [33], [49] and reported preliminary liver shape reconstruction results. However, in the abdominal regions, the available 2D contours or visual cues are poor. We found limitations in learning dense deformation fields and capturing distant features from low-contrast projection images. The improved IGCN+ framework presented herein addresses these problems and exhibits good estimation performance for multiple abdominal organs.

III. Methods

We consider 3D-CT/MRI volumes obtained for pre-treatment planning and unregistered 2D projection images obtained during image-guided therapy. We do not focus on automatic segmentation techniques, and we assume that the organ contours are segmented from planning CT images and organ shapes are modeled as triangular surface meshes as a preprocessing step.

Let M be the initial pre-treatment mesh generated from 3D-CT planning images and I be the 2D projection image obtained from the target state. X-ray or endoscopic images are candidates for I; however, in our experiments for quantitative evaluation of the proposed 2D/3D deformable registration method, DRR images were used. These DRR images were generated from 4D-CT images for performance analysis targeting non-linear motion and shape variability of abdominal organs during respiratory motion.

The left and central images of Fig. 1 show two typical examples of M and I, respectively. In the projection images, the abdominal organs are invisible. Further, the anatomical variability in organ shape and location between patients is apparent. The right images are overlaid with the projected vertices of the mesh, indicating initial misalignment between M and I. The diaphragm shape visualized in the DRR does not match the projected initial shape because the two states differ in terms of patient condition (e.g., posture and respiratory phase). The deformation is nonlinear and exhibits local rotation and sliding motion [45] in 3D, and simple linear transformation is not sufficient to register the two states.

A. IGCN+ Architecture

Fig. 2 illustrates an outline of the IGCN+ architecture and deformable model registration process. The outline of the previous IGCN framework is also illustrated to clarify the difference of the network architecture. The IGCN+ is a generalized, organ-independent framework that integrates an image translation network g and a vertex transformation network f. Various architectures are acceptable for each network model, but we concentrated on verifying the effectiveness of learning the deformation map from the projection image in the 2D/3D deformable registration problem. We hence employed the de facto standard supervised learning models, i.e., a U-Net-based network [50] and graph convolutional network [23] for g and f to match the basic network structure in IGCN+ with that used in IGCN.

g takes a 2-channel image formed by I and a semantic label S generated from M. In our experiments, the input image size was $640 \times 640 \times 2$; however, the method is not limited to a particular size. g learns the generation of a 3-channel displacement map u, which represents a spatial mapping function in 2D space. Then, f receives feature vectors from u and M. M is projected onto the 2D image space, and the pixel values of u, i.e., 3D displacement vectors, are sampled from the projected points p. The 3D vertex coordinates of M and corresponding 3D displacement vectors are concatenated into f for learning deformation. Finally, f generates a deformed mesh registered to I.

The IGCN+ implements a new 2D/3D deformation learning scheme characterized by f and g. Existing projected point sampling methods struggle to capture image features

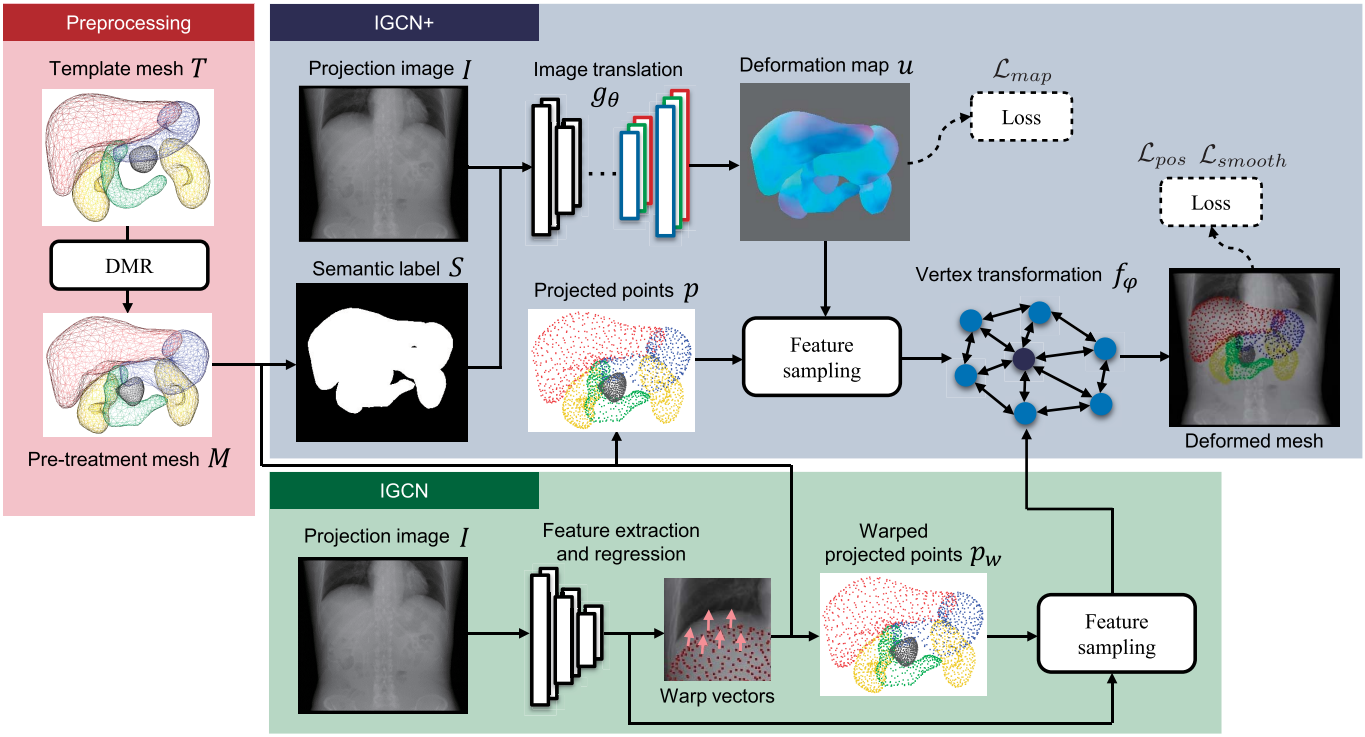

Fig. 2. Full IGCN+. The image translation network g_θ learns to generate a displacement field u. The vertex coordinates and corresponding 3D displacement vectors are concatenated into the GCN f_φ for mesh deformation learning. IGCN relies on both feature extraction and reference point optimization in the CNNs.

distant from the initial mesh [22], [33]. P2M [22] employs CNN-based feature encoding with hierarchical extension to fit an ellipsoid template to various 3D objects. However, it concentrates on mesh deformation and neglects the target object motion. Abdominal organs contain both local deformation and global translation, and the projection images have no clear edges. The displacement map u and composite function $g \circ f$ address the non-linear organ motion and deformation. We describe the roles of these two functions in the next sections.

B. Displacement Mapping Function

CNN-based encoding schemes struggle to learn image features that are distant from the initial template. In our previous IGCN framework [33], we map the projection point to a new position at which a higher probability of obtaining effective image features is expected. As shown in Fig. 2, the CNN part in the IGCN outputs both image features and 2D warp vectors for the projected points obtained from the pre-treatment mesh M. The warped projection points p_w are calculated by adding 2D warp vectors to the projection points p. The image features are sampled from the warped projection points p_w. This scheme improves the registration results; however, when feature encoding and spatially discrete references at the sparse vertices of the mesh are simultaneously updated, the weight optimization for CNNs becomes unstable. In general, in CNNs including self-attention networks, the loss is calculated based on image features obtained from pixels; therefore, the sampling points are fixed to the input image throughout the training. In the IGCN framework, in addition to being spatially sparse, the sampling points are updated dynamically as the 3D mesh deforms at each training epoch. With such sparse sampling, most of the unsampled image features are not used in the loss calculation, and the appropriate gradients needed for continuous updating of the CNN weights are not obtained. When some pixels are stochastically sampled and used for loss calculation in this way, the optimization of the CNN weights becomes unstable, resulting in locally optimal training.

The method newly proposed in this study involves learning of the transformation function g_θ from I to u based on the supervised learning framework via the image generative network defined by the weight parameter θ. Fig. 3 illustrates the g_θ learning process for the liver. Our framework assumes that the initial and target deformed meshes are registered and have the same topology, i.e., the two meshes have the same number of vertices with point-to-point correspondence (Fig. 3(a)). These registered meshes can be obtained from a template mesh of each organ through deformable mesh registration (DMR) [45] in an automated preprocessing process. The 3D displacement vector d_i of each vertex v_i is obtained from the corresponding points before and after deformation (Fig. 3(b)). Next, a 3-channel projection image (Fig. 3(c)) is obtained by transforming d_i from Euclidean to color space; the displacement vector d_i is defined discretely for each vertex of a single organ mesh and transformed to the surface color of the initial mesh. Because the potential application of this framework is radiotherapy, we assume that the relative angle

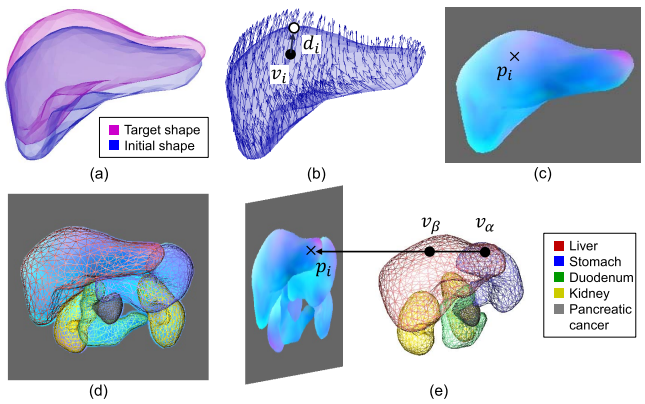

Fig. 3. Learning scheme for image translation function generating spatial mapping function *u*. (a) DMR between initial and target 3D shapes, (b) 3D displacement vectors obtained from corresponding vertices, (c) forward displacement map and sampling, (d) displacement map in abdominal regions overlaid with their meshes, and (e) one-to-many correspondence between displacement field and mesh.

and distance between the patient's body and projection plane is fixed. The surface of the mesh is rendered by locating a camera at the X-ray irradiation position of radiotherapy while considering occlusion using a depth buffer. This means the displacement vectors of the anterior surfaces are projected into 2D space while overwriting those of the posterior surfaces. This problem is resolved in the GCN through embedded learning using 3D vectors obtained from the displacement map as well as the local shape and topologies at each mesh vertex.

The 2D region of the patient-specific organ obtained by projecting the initial mesh can be used as a semantic attention label S. Here, the proposed g_θ defines the transformation:

$$u = g_\theta(I, S). \tag{1}$$

S is used to give alignment information about the initial organ meshes in the projection image domain. It works as an additional attention to the image translation network and contributes to learning the target displacement field that should to be generated in the organ region. Both I and S are treated as the 2-channel input image for stable learning and convergence of the network parameters.

Fig. 3(d) shows the u used for supervised learning and formed from the meshes of the five abdominal organs. Here, u is a projection of the volumetric displacement field that expresses the 3D mesh deformation; thus, the projection points p_i in the map are referenced from the multiple vertices v_α, v_β in the mesh (Fig. 3(e)). In this case, identical displacement vectors can be assigned to all vertices mapped to p. However, v_α, v_β form parts of different organs and different parts of the same organ (e.g., the anterior and posterior); thus, they must each be able to express different displacements. This problem is resolved in the GCN described below through embedded learning using 3D vectors obtained from the displacement map as well as the local shape and topologies at each mesh vertex.

C. Vertex Transformation Function

The vertex transformation function f updates each vertex in the mesh using the generated u and template mesh M structure. Thus, f is responsible for the spatial transformation of each vertex in the mesh based on the GCN, where

$$\hat{v_i} = f_\varphi(v_i, u(p_i)). \tag{2}$$

Here, v_i is the vertex coordinates after normalization; $u(p_i)$ is the 3D displacement vector obtained from the corresponding projection point p_i in the displacement map; f_φ is composed of the GCN and a learnable parameter φ, where the input is a vector having v_i and $u(p_i)$ concatenated; and the output is the predicted value $\hat{v_i}$ of the vertex coordinates. Deformation of the entire mesh is calculated by transforming all vertices $v_i \in \mathcal{V}(i = 1, 2, \ldots, n)$ composing M using the trained function $f_{\hat{\varphi}}$.

One-to-many correspondence between the deformation map $u(p_i)$ and the vertex displacement of the mesh can be addressed because the vertex transformation function f learns the relationship between the feature vector $(v_i, u(p_i))$ and the vertex displacement. For example, when two different vertices v_α and v_β in the initial mesh are projected onto the same projection point p, the GCN can distinguish their features because the vertex transformation function f uses $(v_\alpha, u(p))$ and $(v_\beta, u(p))$ as the input to predict the vertex transformation. This means that the GCN can learn and distinguish deformations for the corresponding anterior and posterior surfaces of the organ based on the initial vertex positions.

For the GCN layers, graph convolution is applied to obtain hierarchical topological features in non-Euclidean space [23]. The mesh is a type of graph $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of edges. The per-vertex features are shared with the neighbor vertices. The GCN employed in this study consisted of eight sequential graph convolutional layers, each of which is defined in Eq. (3).

$$X^{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}X^{(l)}W^{(l)}), \tag{3}$$

where $X^{(l)}$ and $X^{(l+1)}$ denote the feature matrix before and after convolution, respectively. In our experiments, $X^{(l)}$ was the concatenation of the vertex coordinates v_i and displacement vectors $u(p_i)$, and W was the learnable parameter matrix. $A \in \mathbb{R}^{n\times n}$ was the adjacency matrix, i.e., a symmetric matrix with binary values, in which element A_{ij} was 1 if there was an edge between v_i and v_j, and 0 if the two vertices were not connected. $D \in \mathbb{R}^{n\times n}$ was the degree matrix, i.e., a diagonal matrix, in which each element D_{ii} represented the number of edges connected to v_i. The template mesh was deformed by updating $X^{(l)}$.

D. Loss Functions

The parameters (θ, φ) of the overall network are simultaneously updated and optimized by minimizing an objective function. In this section, we introduce three loss functions to achieve 2D/3D deformable mesh registration under the constraint of smooth deformation.

The ground-truth vertex coordinates of the target meshes are obtained from the deformable registration process. To strictly evaluate the point-to-point correspondence, we define the mean distance loss $\mathcal{L}_{pos}$ of the vertex positions between the estimated shape and the ground truth as

$$\mathcal{L}_{pos} = \frac{1}{n}\sum_{i=1}^{n} \|v_i - \hat{v}_i\|_2^2, \tag{4}$$

where $v_i \in \mathcal{V}(i = 1, 2, \ldots, n)$ is the target 3D position, and $\hat{v}_i$ is the predicted position. This loss function induces the convergence of the estimated vertex to the correct position.

In our problem setting, the organ deformation is spatially non-linear and heterogeneous but expected to remain within a limited range. To preserve the curvature and smoothness of the initial surface, we use a regularization loss $\mathcal{L}_{smooth}$ that evaluates a discrete Laplacian of the mesh:

$$\mathcal{L}_{smooth} = \frac{1}{n}\sum_{i=0}^{n} \|L(v_i) - L(\hat{v}_i)\|_2^2, \tag{5}$$

Here, $L(\cdot)$ is the Laplace-Beltrami operator and $L(v_i)$ is the discrete Laplacian of v_i defined by $L(v_i) = \sum_{j \in N(v_i)}(v_i - v_j)/N(v_i)$. $N(v_i)$ is the number of adjacent vertices v_j of the 1-ring connected by the vertex v_i. This loss constrains the shape changes from the initial state and avoids generation of unexpected surface noise and low-quality meshes.

In addition to evaluating the mesh vertex coordinates, accurate prediction of u improves the 2D/3D deformable registration results. Specifically, stable learning of u is important when the target contains both translation and local deformation. Thus, we introduce the displacement map loss $\mathcal{L}_{map}$ determined by the mean absolute error (MAE), such that

$$\mathcal{L}_{map} = \|u - \hat{u}\|_1, \tag{6}$$

where u is the target displacement map and $\hat{u} = g(I, S)$ is the predicted displacement map transformed from I. The existing 2D/3D deformable registration framework [22], [48] does not use a displacement map, and this study is the first to investigate the performance of the newly designed loss function.

The full objective $\mathcal{L}$ is defined as the weighted sum of the above three loss functions:

$$\mathcal{L} = \mathcal{L}_{pos} + \mathcal{L}_{map} + \lambda\mathcal{L}_{smooth}. \tag{7}$$

The loss function values are normalized to $[0, 1]$ using the maximum values in each space. Here, to facilitate supervised learning, we used 0.1 for λ, after examination of several parameter sets. In the next section, we report ablation studies and show the effectiveness of individual loss functions.

The optimized deformable registration model $g_{\theta^*} \circ f_{\varphi^*}$ is obtained by solving

$$g_{\theta^*}, f_{\varphi^*} = \arg\min_{g_\theta, f_\varphi} \mathcal{L}(g_\theta, f_\varphi). \tag{8}$$

These are applied to the developed framework at each epoch to train (g_θ, f_φ).

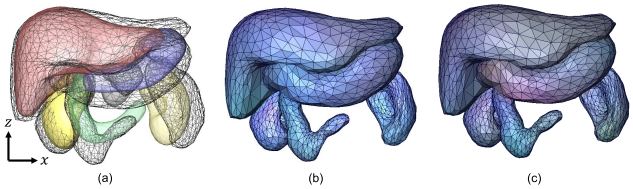

Fig. 4. Statistical models of abdominal organs with respiratory deformation: (a) mean (translucent) and patient-specific (mesh) shapes, and (b) first and (c) second principal components of vertex displacements. The colors represent 3D displacement vectors.

E. Statistical Generative Models

In this section, we introduce a data augmentation method based on statistical generative models to overcome the limited training data volumes. Displacements that reflect the statistical properties of respiratory deformation, as obtained from 4D-CT data, are supplied to a mesh obtained from 3D-CT data. Specifically, for a mesh generated from a 4D-CT volume, DMR [45] is conducted between all cases to obtain a mesh with the same topology. Then, principal component analysis is conducted to obtain a statistical model of the shape and displacement.

Fig. 4(a) shows patient-specific shapes obtained via 4D-CT and the average shapes calculated from data for multiple patients. Figs. 4(b) and (c) show results obtained by transforming the first and second principal components, respectively, of the displacement to the RGB space and visualizing these data as color maps for the mesh surfaces. The displacement z-component is large because of the characteristics of respiratory motion; however, the local displacement distributions of each organ have different orientations and sizes. The statistical d_i is defined as

$$d_i = \sum_{k=0}^{m} \omega_k c_i^{(k)}, \tag{9}$$

where $c_i^{(0)}$ is the mean displacement at vertex v_i and $c_i^{(k)}(k \geq 1)$ is the kth component of the displacement. Further, ω_k is the weight parameter for each component, and can be changed to yield various d_i values and express the statistical deformation of the 4D-CT data.

Augmented data for supervised learning can be obtained by deforming the registered mesh M obtained from the 3D-CT volume based on d_i. In other words, the vertex coordinates are updated for each v_i of M as $v_i \leftarrow v_i - d_i$. The set of the deformed mesh M_d and the projection image I obtained from 3D-CT volume is used as the input data. The pre-update mesh M can be used as the target shape of the true value corresponding to I. Network $g \circ f$ training is implemented by randomly changing ω_k for each epoch and generating augmented data with various deformation variations online.

IV. Experiments

To evaluate the performance of the proposed method and its potential application in clinical settings, we conducted the following three experiments: 1) a comparison with conventional 2D/3D deformable registration methods, 2) deformation

prediction for multiple organs, assuming a clinical application of moving-target tracking in radiation therapy, and 3) an ablation study to demonstrate the effectiveness of individual components. We implemented our methods using Python 3.9 and TFlearn with a TensorFlow background. We used 1 for each training batch, 200 for the total number of training epochs, and the ADAM optimizer with a learning rate of 1×10^{-4}. Our code and demonstration movies are available online at https://github.com/meguminakao/IGCN.

A. Dataset

3D-CT volumes of 124 cases and 4D-CT volumes of 35 cases were acquired from various patients who underwent intensity-modulated radiotherapy in Kyoto University Hospital. This study was performed in accordance with the Declaration of Helsinki and was approved by our institutional review board (approval number: R1446). Each 4D-CT volume consisted of 10 time phases (t =0, 10, $\cdots$, 90%) for one respiratory cycle and was measured under respiratory synchronization, with $t = 0$ and $t = 50$ corresponding to the end-inhalation and end-exhalation phases, respectively. Thus, 474 3D-CT volumes were used.

Each 3D-CT volume consisted of 512 $\times$ 512 pixels and 88-152 slices (voxel resolution: 1.0 mm $\times$ 1.0 mm $\times$ 2.5 mm). During routine clinical procedures, the following regions were labeled by board-certified radiation oncologists: the entire body, stomach, liver, duodenum, left and right kidneys, and pancreatic GTV. We generated surface meshes (400-500 vertices and 796-996 triangles for one organ) from the region labels and obtained organ mesh models with point-to-point correspondence using DMR. The DMR algorithm and the registration performance for the abdominal organ shapes were reported previously [45], and template meshes registered to patient-specific organ shapes with a 0.2 mm mean distance (MD) error and 1.1 mm Hausdorff distance (HD) error, on average, were confirmed. This registration error was sufficiently small for the use of ground-truth meshes.

We generated DRRs from 3D-CT and 4D-CT volumes. The number of 4D-CT cases were limited; therefore, we adopted a 3-fold cross validation, which divided 35 patients into three groups of 12, 12, and 11. Since one 4D-CT data consist of ten-frame sequential volumes, the total number of test data was 350. In addition, as statistical data augmentation, we calculated the mean and Eigen displacement from 4D-CT data for 23 patients (i.e., 230 training volumes), excluding the test data; these were then adopted to the 3D-CT volumes of 124 patients for learning while continuously generating variations of the organ displacement associated with respiration. The weight parameters were determined as $(\omega_0, \omega_1, \omega_2) = (2, 1, 0)$ after examination of the prediction performance with several parameter sets. We summarized the selection method and the effect of the data augmentation on the registration performance in Section IV-D.

B. Method Comparison

The first experiment was designed to quantitatively and qualitatively compare the registration results of the proposed and existing methods. 3D shapes of the liver, stomach, duodenum, left/right kidneys, and pancreatic cancer GTV were used as the reconstruction targets, and the registration errors were compared. The liver was in contact with the diaphragm, and the upper 2D contour was detectable, but the contours on the lateral and lower regions could not be visually confirmed. The contours of the other four abdominal organs (stomach, duodenum, kidney, and pancreatic cancer GTV) could not be visually confirmed on the 2D projection image, and 3D shape reconstruction was even more challenging. In this experiment, these five organs were used as the estimation target for the performance evaluation and error analysis.

1) Baseline and Experimental Conditions: Few existing methods can achieve 2D/3D deformable model registration from a single-viewpoint projection image for deformable organs. We selected the following four existing learning methods for comparison; two designed for 3D medical images where any graph structures are not used, and the other two that employ image-to-graph convolutional architectures. We then compared their 2D/3D registration performance with that of the proposed IGCN+.

AF: an image-based registration method that obtains a 2D affine transformation between the initial and target projected images and estimates the 3D global affine transformation for the initial template mesh using a linear regression. This model has been widely used for 2D/3D rigid registration [6], [7], and was added to confirm the complexity of our deformable registration problem.

VM: a non-rigid image-based registration approach using VoxelMorph [13], which calculates the 2D dense displacement field between the initial and target projected images and estimates the 3D displacement for each vertex. This model does not use any graph structures and learns the 3D displacement from the initial position v_i and the corresponding 2D displacement obtained from projection point p_i. This transformation is the same as Eq. (2), but performed using a simple regression model.

P2M: an image-to-graph network model designed for general images [22]. In our study, the ground-truth position is obtained for each vertex; thus, the Chamfer loss used in P2M was changed to the $\mathcal{L}_{pos}$ defined in Eq. (4), and the remaining loss function was used without alteration. Hierarchical learning was not applied so that the prediction process would match those of the other methods.

IGCN: our previous framework [33]. Its basic structure is similar to that of the P2M model, but implements the additional learning scheme for warped projection to obtain better image features for registration. This model does not learn the dense 3D displacement field, and this point is the main difference between it and the proposed framework.

We evaluated the 3D shape and position accuracies for the predicted organs using three error indices, mean distance (MD), the Hausdorff distance (HD) [53], and MAE between surfaces. The Dice similarity coefficient (DSC) was also used. We obtained a mesh with vertex correspondences by applying DMR for each organ [45] and used it as the target shape with ground-truth coordinates. The MD and HD are the average and maximum values, respectively, of the bidirectional distance

defined by the two nearest vertices between the predicted and ground-truth mesh; these values quantify the error between shapes. The MAE is the average Euclidean distance between the predicted and correct positions of the corresponding vertex and reflects the prediction error for each vertex. The DSC quantifies the overlap between the 3D regions of two meshes; a higher value indicates better prediction performance.

Here, two approaches, referred to as single and multiple reconstruction, can be considered to estimate the shapes of multiple organs: single reconstruction learns each individual organ and multiple reconstruction simultaneously learns multiple organs as a tetrahedral mesh by generating the connectivity between organs. Multiple reconstruction increases both the number of vertices in the mesh to be estimated and the shape expression complexity, but it can effectively learn positional relationships between organs and the deformation interactions. In the method comparison, we summarize the results of multiple reconstructions to fairly compare the image-based methods and image-to-graph framework. Thus, the shape error was calculated for each organ after simultaneously predicting the 3D shapes of all five organs from one DRR. We further analyze the difference between the two reconstruction approaches in the next section.

Verification was conducted using the following two conditions with respect to initial alignment of the template mesh: the "w/o setup error" condition, which used the initial position in the first phase ($t = 0$); and the "w/ setup error" condition, for which the initial position was set as the position translated in 3D using random noise, with the maximum displacement being twice the average respiratory displacement. We considered the "w/ setup error" condition because of differences in the setup of input images between the experiment and clinical situations. In the "w/o setup error" condition, the initial alignment of DRR images is determined using the 4D-CT end-inhalation phase. This means the relative position of the patient to the bed during radiotherapy is given. However, as such a strict setup of the patient's body is probably difficult to achieve in a clinical situation, we considered that the possibility of operational misalignment of the relative positions between the patient and the bed should be taken into account. To reproduce this, we analyzed the prediction performance with noise added to the initial alignment. The maximum range of the added noise was 17.0 mm in each direction, which was sufficiently larger than the real setup error for bony structures and variation in the 3D positions of pancreatic tumors due to respiration [51]. For both conditions, the organ shape and position were set as unknown for all phases, dynamic properties and hysteresis caused by time changes were neglected, and this problem was regarded as a problem of static reconstruction of the organ shape in each phase. For each method, training was performed through data augmentation using the statistical generative model mentioned above.

2) Comparison of Results With Baseline: Table I lists the average values and standard deviations of the evaluation indices obtained for 350 test data points for each organ. Here, "Initial" refers to the difference between the known 3D shape of the first phase $t = 0$ and the target state of the nine phases $t = 10, 20, \ldots, 90$. VM and IGCN performed slightly better than the other methods, and the proposed IGCN+ outperformed the other methods for all organ. There is a relatively large difference between the values obtained by IGCN and IGCN+, and these differences increased for the stomach, duodenum, and pancreatic GTV. Interpatient shape variation in these organs or regions is relatively large [45], and their contours are not visible on the 2D projection image. This implies that learning the pixel-level deformation map is effective for the 2D/3D registration of soft organs with shape variability.

To further confirm the robustness of the methods, we accumulated error indices for all five organs and the pancreatic GTV and compared the registration performance with respect

TABLE I
QUANTITATIVE COMPARISON OF MULTI-ORGAN SHAPE RECONSTRUCTION. MEAN ± STANDARD DEVIATION OF MAE, MD, HD AND DSC

Liver	Shape reconstruction errors			
	MAE [mm]	MD [mm]	HD [mm]	DSC [%]
Initial	8.93 ± 4.68	3.75 ± 2.31	14.38 ± 6.77	89.45 ± 6.67
AF	8.66 ± 5.20	3.38 ± 2.69	13.27 ± 6.74	90.79 ± 6.39
VM	7.59 ± 3.18	3.03 ± 1.37	12.32 ± 5.20	91.58 ± 3.89
P2M	8.38 ± 2.55	3.63 ± 1.29	12.90 ± 3.91	89.95 ± 3.82
IGCN	6.78 ± 2.16	2.65 ± 0.78	12.26 ± 3.51	92.89 ± 2.40
IGCN+	**6.09 ± 2.27**	**2.14 ± 0.84**	**10.27 ± 4.02**	**94.48 ± 2.22**

Stomach	Shape reconstruction errors			
	MAE [mm]	MD [mm]	HD [mm]	DSC [%]
Initial	8.27 ± 4.58	3.23 ± 2.04	11.23 ± 6.27	81.36 ± 12.42
AF	7.31 ± 3.72	2.79 ± 1.60	9.63 ± 4.74	84.12 ± 9.69
VM	7.08 ± 2.77	2.74 ± 1.04	9.62 ± 4.37	84.47 ± 6.64
P2M	7.52 ± 2.52	3.06 ± 0.99	9.85 ± 3.36	82.36 ± 6.65
IGCN	6.33 ± 2.83	2.26 ± 0.82	9.15 ± 3.57	86.86 ± 5.77
IGCN+	**5.94 ± 2.95**	**1.98 ± 0.88**	**7.88 ± 3.76**	**89.18 ± 5.48**

Duodenum	Shape reconstruction errors			
	MAE [mm]	MD [mm]	HD [mm]	DSC [%]
Initial	7.95 ± 4.40	2.96 ± 1.93	11.05 ± 5.85	74.73 ± 17.04
AF	7.23 ± 4.74	2.73 ± 1.98	10.23 ± 6.07	76.58 ± 13.66
VM	7.13 ± 2.80	2.61 ± 1.13	10.13 ± 4.72	77.42 ± 10.84
P2M	7.69 ± 2.42	2.94 ± 0.97	11.04 ± 4.50	73.77 ± 9.43
IGCN	7.15 ± 2.17	2.58 ± 0.70	10.68 ± 4.05	77.02 ± 7.37
IGCN+	**5.84 ± 2.44**	**1.87 ± 0.75**	**8.83 ± 4.50**	**83.82 ± 7.60**

Left kidney	Shape reconstruction errors			
	MAE [mm]	MD [mm]	HD [mm]	DSC [%]
Initial	8.27 ± 4.79	3.13 ± 1.76	10.81 ± 5.43	84.65 ± 9.15
AF	7.09 ± 5.08	2.87 ± 1.85	9.87 ± 5.78	85.79 ± 8.24
VM	6.32 ± 3.11	2.39 ± 1.12	8.82 ± 4.11	88.17 ± 5.93
P2M	7.24 ± 2.56	2.72 ± 0.97	9.64 ± 3.61	86.50 ± 5.26
IGCN	6.51 ± 1.89	2.37 ± 0.71	9.99 ± 3.31	87.83 ± 4.22
IGCN+	**5.44 ± 1.98**	**2.07 ± 0.81**	**8.42 ± 3.46**	**89.83 ± 4.52**

Right kidney	Shape reconstruction errors			
	MAE [mm]	MD [mm]	HD [mm]	DSC [%]
Initial	9.44 ± 5.94	3.43 ± 1.99	11.67 ± 6.48	83.80 ± 9.59
AF	8.39 ± 6.11	3.27 ± 2.32	10.85 ± 6.77	84.30 ± 9.20
VM	7.29 ± 4.48	2.56 ± 1.29	9.42 ± 5.28	87.75 ± 6.43
P2M	8.52 ± 3.56	3.11 ± 1.16	10.66 ± 4.52	84.79 ± 5.92
IGCN	7.79 ± 3.07	2.70 ± 0.81	10.56 ± 4.02	86.45 ± 4.52
IGCN+	**5.86 ± 3.10**	**2.05 ± 0.80**	**8.35 ± 4.24**	**90.26 ± 3.93**

Pancreatic GTV	Shape reconstruction errors			
	MAE [mm]	MD [mm]	HD [mm]	DSC [%]
Initial	5.20 ± 3.56	2.20 ± 1.66	6.14 ± 3.74	81.19 ± 13.70
AF	4.61 ± 2.60	2.00 ± 1.42	5.58 ± 3.08	82.75 ± 11.68
VM	5.10 ± 2.22	2.17 ± 1.06	6.03 ± 2.42	80.54 ± 10.14
P2M	5.56 ± 2.11	2.43 ± 1.06	6.55 ± 2.18	78.31 ± 10.33
IGCN	4.82 ± 1.77	2.01 ± 0.78	6.80 ± 1.95	80.00 ± 8.53
IGCN+	**3.54 ± 1.84**	**1.30 ± 0.69**	**4.50 ± 2.16**	**88.48 ± 6.43**

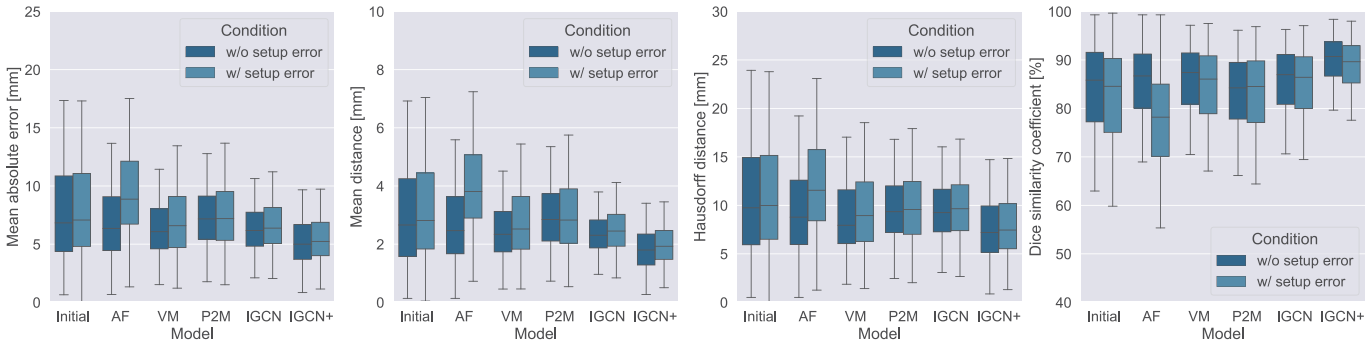

Fig. 5. Shape registration errors for abdominal organs with respect to "w/o setup error" and "w/ setup error" conditions. Each box plot presents accumulated evaluation indices for 350 test data obtained from registered meshes of the liver, stomach, duodenum, left/right kidneys, and pancreatic GTV.

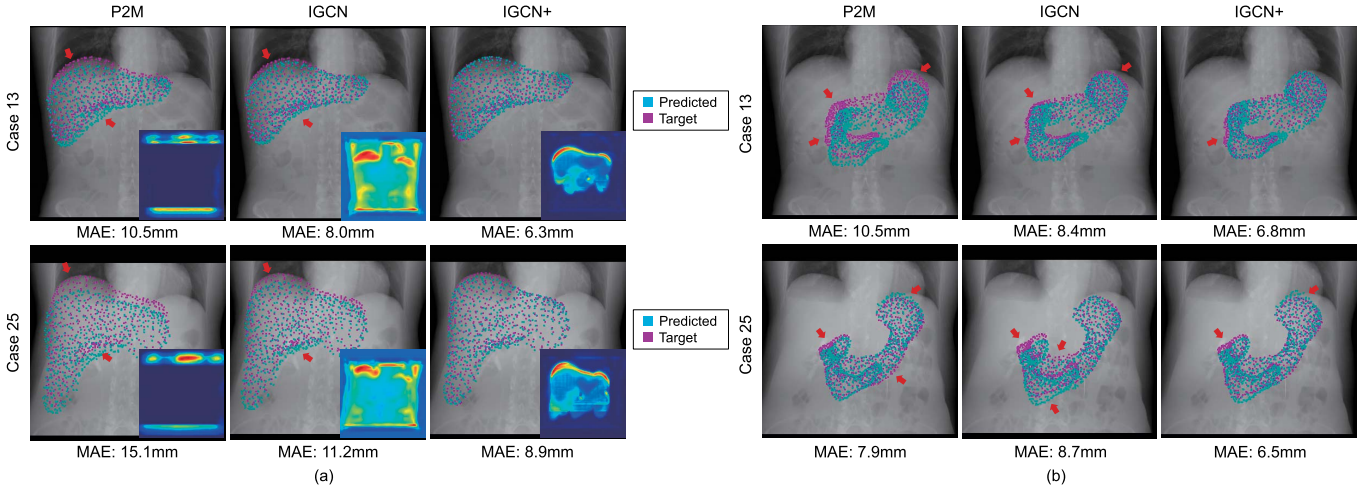

Fig. 6. Visual comparison of methods with respect to the shape reconstruction of abdominal organs for average (Case 13) and maximum (Case 25) error cases. (a) Predicted (cyan), target shapes (magenta) with latent image features (heat map) for the liver and (b) predicted results for the stomach and duodenum.

to the "w/o setup error" and "w/ setup error" conditions described above. Fig. 5 presents the box plots of the respective errors for the two conditions. For the MAE and MD values, the initial errors increased by 5.4% (0.43 mm) and 8.3% (0.24 mm), but the increases in the errors of the proposed IGCN+ were only 4.4% (0.24 mm) and 6.3% (0.11 mm), respectively. Thus, stable prediction could be achieved even with setup error in the patient posture or differences in the initial conditions associated with the 3D shape alignment. The errors in AF and VF are more affected by the added noise, suggesting that graph convolution can improve registration stability in initial value problems. As for both conditions, significant differences (one-way analysis of variance, ANOVA; $p < 0.05$) were confirmed for the conventional methods for all indices.

The smoothness of the predicted shape and the mesh quality could not be evaluated using the above error indices only; therefore, we qualitatively confirmed the estimation results by visualizing the estimated shape. Fig. 6(a) shows results obtained by superimposing the liver shape predicted through DRR of the end-exhalation phase ($t = 50$) for each method, for the case in which predictions were based on the mean shape error (Case 13) and the case for which the shape error was largest (Case 25). Fig. 6(b) visualizes the estimated stomach and duodenum shapes of the same cases. Magenta indicates the true liver and position, and cyan shows the predicted shape. Arrows indicate relatively large shape errors. A heat map on the right-bottom of each figure in Fig. 6(a) show the sum of the latent image features in the same feature encoding layer. We note that the 3D shapes of five abdominal organs are predicted simultaneously from one DRR, and liver, stomach and duodenum are selected in each figure for clarity of visual comparison.

The proposed method predicted a deformation that is similar to the target 3D shape despite the fact that visual confirmation of the contours was not achieved for many liver areas, and only extremely low-contrast textures were visible. Visual comparisons of the latent image features and prediction results reveal that P2M responded strongly to the boundary of the field of view, with large errors at locations with low correlation with the body contour movements, as indicated by the arrows in Fig. 6(a). Cases in which the prediction can fail with large displacement, even if the edge around the diaphragm is relatively clear, are shown. For IGCN, responses to the low-contrast edges and texture are apparent. However, the errors increase in the lower liver, where the edge could not be

TABLE II
COMPARISON OF DEFORMABLE REGISTRATION PERFORMANCE FOR SINGLE AND MULTIPLE RECONSTRUCTIONS

	Single reconstruction			Multiple reconstruction		
	MD [mm]	HD [mm]	DSC [%]	MD [mm]	HD [mm]	DSC [%]
Liver	2.07 ± 0.76	9.96 ± 3.98	94.65 ± 2.02	2.12 ± 0.84	10.27 ± 4.02	94.48 ± 2.22
Stomach	3.64 ± 1.84	12.57 ± 8.53	79.76 ± 9.29	1.98 ± 0.88	7.88 ± 3.76	89.18 ± 5.48
Duodenum	3.23 ± 1.35	12.08 ± 4.84	72.51 ± 13.23	1.87 ± 0.75	8.83 ± 4.50	83.82 ± 7.60
Left kidney	3.30 ± 1.69	11.38 ± 4.86	83.43 ± 8.95	2.07 ± 0.81	8.42 ± 3.46	89.83 ± 4.52
Right kidney	3.19 ± 1.46	11.85 ± 5.28	83.74 ± 8.31	2.05 ± 0.80	8.35 ± 4.24	90.26 ± 3.93
GTV	2.34 ± 1.42	6.85 ± 2.94	79.07 ± 12.91	1.30 ± 0.69	4.50 ± 2.16	88.48 ± 6.43

visually confirmed. The proposed IGCN+ generated features for the abdominal organ areas and their surroundings, which yielded good predictions for the lower area of the liver. The estimation performance of each method for the stomach and duodenum tended to be similar. Because they are smaller than the liver, the median error is relatively smaller, but the error variability is conversely larger, with misalignments and local shape errors depending on the case.

C. Multiple-Organ Deformation Prediction

In the second experiment, we aimed to verify the organ deformation and displacement prediction performance assuming clinical applications to tumor tracking radiation therapy. We verified whether the final performance achieved the 3D organ area identification accuracy required for real-time localization of pancreatic GTV and surrounding OAR volumes. We also confirmed effectiveness of multiple reconstruction for 2D/3D deformable registration with comparison to single organ reconstruction. For single reconstruction, we calculated the error by predicting each 3D shape from one DRR for the liver, stomach, duodenum, and pancreatic cancer GTV based on the individually-trained network. For multiple reconstruction, same as the previous experiment, we calculated the shape error for each organ after simultaneous prediction of five abdominal organs.

1) Performance Analysis and Motion Dynamics: Fig. 7 shows the liver, stomach, duodenum and GTV displacements for each phase, as well as the shape reconstruction errors due to single and multiple reconstruction. The mean value for all corresponding vertices was visualized as the centerline, and the standard deviation was depicted as a colored band. Table II lists the errors for each organ with regard to nine frame sequential images ($t = 10, 20, \ldots, 90$) for both single and multiple reconstruction. For the liver, no significant differences between the two approaches were apparent, but significant differences (ANOVA; $p < 0.05$) were confirmed between the two methods for the stomach, duodenum, kidneys and GTV. Shape error improvements of 45.6%, 42.1%, 37.3%, 35.7% and 44.4%, respectively, were obtained for multiple reconstruction. Regarding the quantitative metric tolerance in the American Association of Physicists in Medicine guideline for image registration and fusion [52], MD is 2 - 3 mm and DSC is 80 - 90%. The obtained results show that shape reconstruction can be achieved with accuracy equal to or exceeding this level and, thus, the proposed method is clinically applicable. We measured the computation time for the whole registration process. The average computation time was 48.9 ms (20.4 frames per second), demonstrating the real-time registration performance of the IGCN+.

2) Abdominal-Organ Shape Reconstruction: Cases 13 and 25 are shown as typical shape reconstruction examples in Fig. 8; these cases show the average and maximum shape errors, respectively, for $t = 50$, which had the largest displacement. The quantitative values inserted at the bottom of the figure are accumulated error indices for all five organs and the pancreatic GTV. Fig. 8(a) shows the multiple reconstruction results, and the central image (predicted) shows the vertices of the 3D organ mesh obtained for the input DRR image, where coloring and superimposed visualization were conducted for each organ. The image on the right was obtained by projecting the true (magenta) and predicted (cyan) shapes on the projected image; the shape errors of each organ could be locally confirmed. Unlike with the liver, the contours could not be visually confirmed on the DRR images for the stomach, duodenum, and GTV; however, shape reconstruction with minimal deviation from the ground-truth shape was achieved. The supplemental movie (available online at https://github.com/meguminakao/IGCN) demonstrates the results for 10-frame sequential images with more examples.

Fig. 8(b) shows the results obtained by visualizing the 3D meshes of the abdominal organs using single and multiple reconstruction from two different directions. For single reconstruction, large deviations in the stomach and duodenum positions were noted. Thus, shape reconstruction using only the image features obtained from the 2D area in the DRR image corresponding to each organ was difficult. For multiple reconstruction, good matching was found between the true and predicted shapes. Note that stomach shapes varied considerably between patients because of the stomach contents, and some deviations were observed. Fig. 8 (c) presents other multi-organ shape reconstruction examples with relatively large errors confirmed in the liver and stomach. Despite the fact that very low-contrast textures and the appearance of the projection images differ between patients, the shapes can be stably reconstructed with acceptable errors.

D. Ablation Study

In the third experiment, we conducted an ablation study to analyze the effect of individual components in the developed IGCN+ framework. IGCN+ consists of two networks, and the roles of the semantic label input, graph convolutions, loss functions, and statistical data augmentation were investigated. Because it is difficult to validate all combinations of these factors, we obtained the registration performance when learning

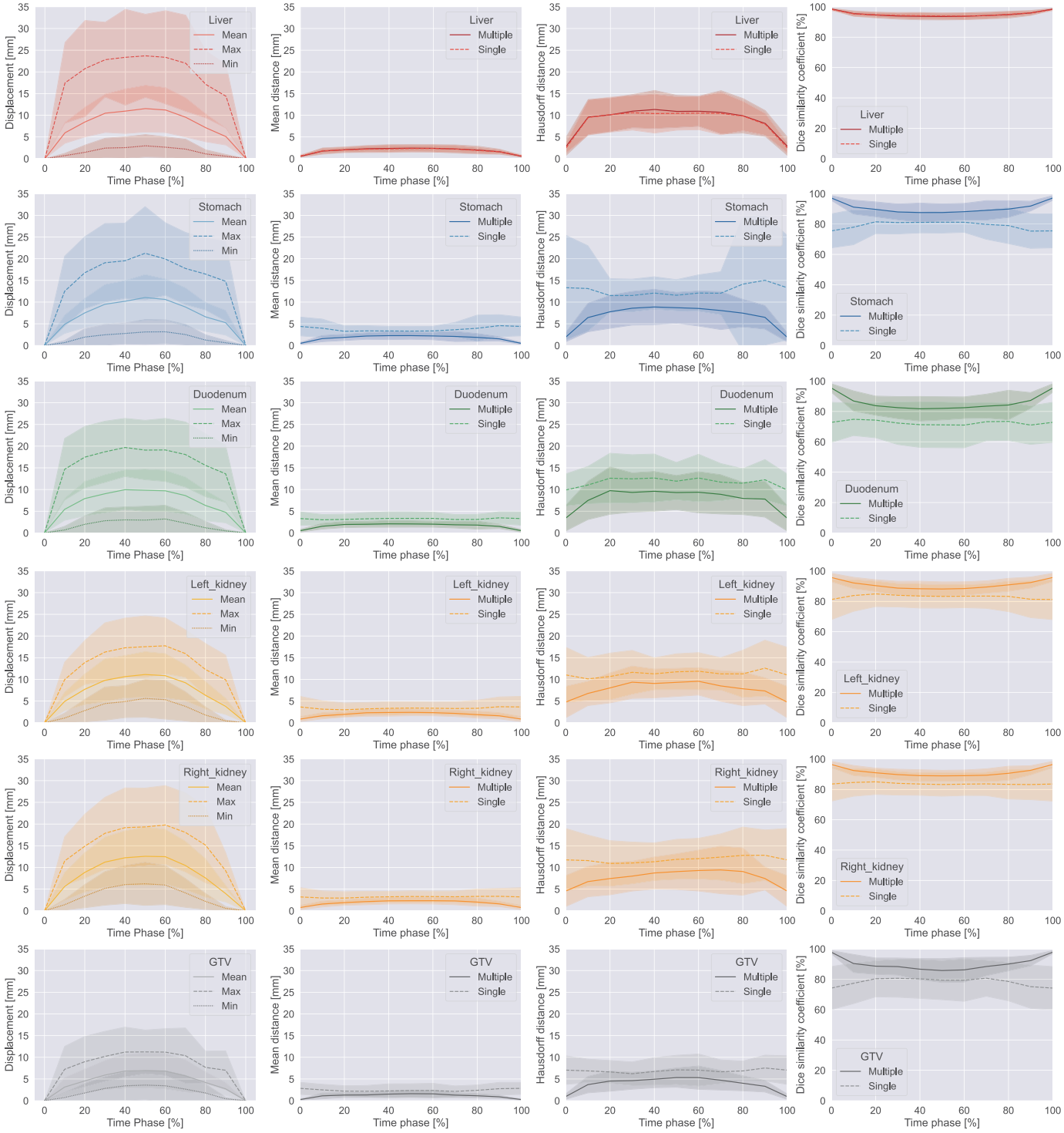

Fig. 7. Displacements and shape reconstruction errors (mean distance, the Hausdorff distance and the Dice similarity coefficient) for five abdominal organs and pancreatic cancer GTV due to single and multiple reconstruction. The mean value for all corresponding vertices are plotted as the centerline, and the standard deviation is visualized as a colored band.

and prediction were performed by excluding each component from the IGCN+. MD, HD, and DSC were employed as the error metrics, and the sum of the errors for all organs were used to simplify the comparison of each model, as shown in the evaluation of all abdominal organs in Fig. 5.

1) Input Images, Graph Convolutions and Loss Functions: Table III shows the effects of the key components in the IGCN+ framework. In the "w/o semantic label" model, semantic label S was excluded, and a 1-channel DRR image was used as the input. In the "w/o graph convolution" model, the vertex transformation function was implemented by a simple regression model, meaning the vertex position was directly estimated from 3D displacement vector $u(p_i)$ without using mesh connectivity. In "w/o displacement mapping loss" and "w/o regularization loss" models, we simply excluded $\mathcal{L}_{map}$ and $\mathcal{L}_{smooth}$ from the full objective $\mathcal{L}$, respectively.

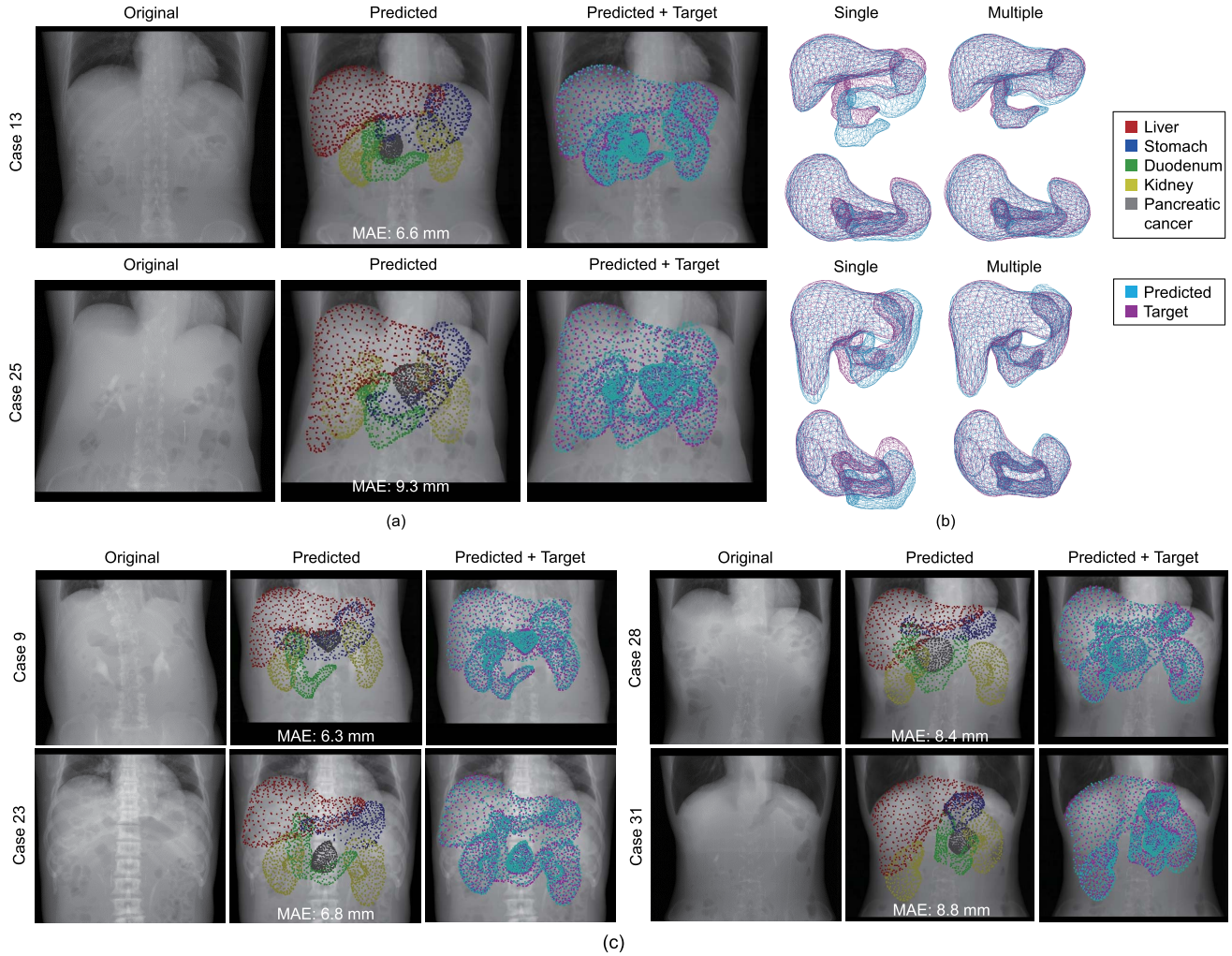

Fig. 8. Shape reconstruction examples and method comparison. (a) Average and maximum error cases, (b) estimated shapes obtained from single and multiple reconstruction, and (c) other examples. The graph convolutions embed per-vertex displacement vector between organs, which results in better estimation performance of the regions with no visual cues.

TABLE III
EFFECT OF KEY COMPONENTS IN IGCN+ ON 2D/3D DEFORMABLE REGISTRATION

Models	Shape reconstruction errors		
	MD [mm]	HD [mm]	DSC [%]
IGCN+	**1.90 ± 0.84**	**8.04 ± 4.16**	**89.34 ± 6.18**
w/o graph convolution	2.45 ± 0.83	8.98 ± 3.89	86.11 ± 6.62
w/o semantic label	3.01 ± 1.14	10.62 ± 4.33	82.67 ± 8.61
w/o displacement map loss	2.26 ± 0.96	9.11 ± 4.44	87.35 ± 7.03
w/o regularization loss	2.03 ± 0.82	8.29 ± 4.08	88.49 ± 6.18

These results show that each component improves registration accuracy. In particular, the semantic label and graph convolution play key roles in our problem settings. We note that the semantic label determines the initial alignment of the template to the projection plane, and graph convolution yields vertex connectivity with the surrounding organs. This result demonstrates the importance of considering these two factors in 2D/3D registration. In addition, both loss functions help reduce the registration error. We also performed additional experiments with L1 loss for $\mathcal{L}_{pos}$ and/or $\mathcal{L}_{smooth}$, which resulted in lower performance compared with the results with L2 loss. In our model, the L2 loss was considered sufficient because there was no sparsity in the displacement vectors to be predicted.

2) Data Augmentation Using Statistical Generative Model: Lastly, we investigated the effects of the different data augmentation methods on 3D organ shape reconstruction considering respiratory displacement variation. First, to determine the $(\omega_0, \omega_1, \omega_2)$ of the statistical generative model, training was conducted using eight parameter sets for the 3D-CT models of 124 cases while generating variations in respiratory motion. The test data and target organs were the same as those described in the Section IV-A. We predicted the shape of the abdominal organs included in the 4D-CT data and calculated the error metrics.

Table IV presents the parameter set and shape errors investigated in order. When only ω_0 was changed, relatively good performance was obtained for $\omega_0 = 2$, which corresponds to

TABLE IV
COMPARISON OF AUGMENTATION METHODS AND THE EFFECT OF WEIGHT PARAMETERS IN THE STATISTICAL GENERATIVE MODEL

Augmentation methods	Shape reconstruction errors		
	MD [mm]	HD [mm]	DSC [%]
Statistical translation			
(1, 0, 0)	2.02 ± 1.05	8.25 ± 4.51	88.66 ± 7.51
(2, 0, 0)	1.93 ± 0.86	8.05 ± 4.16	89.04 ± 6.56
(3, 0, 0)	1.97 ± 0.90	8.40 ± 4.37	88.92 ± 6.37
(1, 1, 0)	1.95 ± 1.02	8.12 ± 4.50	89.07 ± 7.24
(2, 1, 0)	**1.90 ± 0.84**	**8.04 ± 4.16**	**89.34 ± 6.18**
(2, 2, 0)	2.07 ± 0.87	8.45 ± 4.07	88.03 ± 6.97
(1, 1, 1)	2.06 ± 0.90	8.33 ± 4.24	88.31 ± 6.83
(2, 1, 1)	2.04 ± 0.85	8.42 ± 4.06	88.45 ± 6.22
Random translation	2.45 ± 1.13	9.28 ± 4.63	86.18 ± 8.18
w/o augmentation	2.21 ± 1.07	8.76 ± 4.61	87.51 ± 7.82

twice the mean respiratory displacement. When both ω_0 and ω_1 were changed, performance improved when the first principal component was considered but deteriorated when the second one was considered. Thus, we adopted $(\omega_0, \omega_1, \omega_2) = (2, 1, 0)$, which yielded the best performance.

We also compared the following cases: no data augmentation ("w/o augmentation"); random translation, which has traditionally been used as a data augmentation standard; and data augmentation using the proposed statistical generative model. For random translation, twice the mean respiratory displacement was randomly applied in 3D regardless of direction. The statistical generative model was trained to output a direction-dependent displacement based on a displacement vector corresponding to the average respiratory displacement vector and the first principal component for the weights obtained as described above. Table IV shows that the best training performance was achieved using the statistical generative model, which decreased the MD values obtained by "w/o augmentation" and random translation by 14.0% and 22.4%, respectively.

V. DISCUSSION

This study presents a new framework that integrates an image generative network and GCN, which achieves model-based deformable registration for 2D projected images. Unlike image-based 2D/3D registration, a mesh that explicitly defines the organ areas to be estimated can be output. A wide range of clinical applications are possible, such as localization of GTVs and OAR volumes in radiation therapy, and tumor position identification for endoscopic camera images during surgery. The main clinical scenario is to apply the developed framework to markerless tumor-tracking radiotherapy. Abdominal organs such as the stomach, duodenum, and pancreas neighbor each other, but the pancreas cannot be clearly detected, even on 3D CBCT images. Our experiments suggest the possibility of real-time tumor localization from time-series X-ray images or surrogate motion during the actual intervention.

Conventional CNN-based feature encoding in IGCN relied both feature extraction and reference point optimization on CNNs; this problem was resolved through displacement map generation by the image generative network. Additionally, significantly improved estimation accuracy for the stomach, duodenum, and pancreatic cancer GTV were confirmed through simultaneous reconstruction of multiple organs, even when there were no visual cues in the projected images. This outcome is thought to be due to successful localization through convolution of the features of adjacent vertices in the GCN. The experiments showed that learning both a wide range of image features and the positional relationship between organs could improve registration accuracy and prediction stability. Thus, multiple-organ registration works better than single-organ registration, but the performance improvement depends on the measurement condition and fields of view of X-ray images obtained in clinical situations. Analysis of the performance obtained when applied to the limited fields-of-view of clinical X-ray imaging will be required.

To facilitate supervised learning using dense deformation fields, we aimed to address the problem of defining the ground-truth displacement field. In the abdominal region in particular, there may be spatially discontinuous displacements due to the interaction of multiple organs and sliding motion, making it difficult to obtain an accurate 3D displacement field in advance. In contrast, the 2D displacement field generation process in this study is built upon an accurate deformable mesh registration with a Hausdorff distance error of 1–2 mm for each organ, and it has the advantage of generating a 2D displacement field in the projection plane while being able to handle spatially discontinuous displacements between organs. This process could also be further extended to 3D displacement field generation, but the logic for this is currently under consideration because it involves the new problem of defining ground-truth 3D displacement fields for supervised learning.

Our experiments had certain limitations. For example, performance evaluation was conducted using only DRRs as projected images, and further evaluations for X-ray images measured during treatment are required. However, multiple studies have reported that DRR-based learning is effective for prediction from measured X-ray images [4], [19], [20]. The organ shape contained in the 3D-CT data used to generate the DRR could potentially be taken as the true value, yielding a quantitative and highly reliable performance comparison, although the estimation error would increase because of the differences between the X-ray and DRR images. Because the potential application of this framework is radiotherapy, we assume that the relative angle and distance between the patient's body and projection plane is fixed. The surface of the mesh is rendered by locating a camera at the X-ray irradiation position of radiotherapy. Setup error was considered in the performance analysis. This setting is probably a limitation in other applications or modalities, but our framework can generate DRR images by changing the projection angle and can learn the relationships between more projection patterns and the underlying deformations. Applying our framework to other clinical applications and its performance analysis will be our future work.

Our framework requires some preprocessing of 3D organ meshes when building a new database for supervised learning. However, the mesh-based 2D/3D registration uses only

geometrical information (i.e., organ contour or shape) and does not depend on the pixel characteristics of the imaging modality (such as CT, MRI, and CBCT). Therefore, the shape reconstruction performance is independent of particular pre-treatment 3D images and the trained network can be transferred to a newly prepared geometrical dataset. Our future work includes improving the accuracy, applying IGCN+ framework to other clinical application, and further performance analysis with the use of high-resolution X-ray images and other image modalities available in actual treatment settings.

VI. Conclusion

In this study, we proposed an extended image-to-graph convolutional network (IGCN+) that achieves deformable model registration of a 3D organ model for a single-viewpoint 2D projection image. We targeted abdominal organs for multiple shape reconstruction from low-contrast projection images, and verified that clinically acceptable registration accuracy was achieved with a mean distance of less than 2 mm for the stomach, duodenum and pancreatic GTV. The proposed technique could be directly applied for localization of radiation targets and organ-at-risk volumes in radiation therapy, and it could also be applied to a wide range of image-guided interventions.

References

[1] B. Rigaud *et al.*, "Statistical shape model to generate a planning library for cervical adaptive radiotherapy," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 406–416, Feb. 2019.

[2] J. Tokuno, T. F. Chen-Yoshikawa, M. Nakao, T. Matsuda, and H. Date, "Resection process map: A novel dynamic simulation system for pulmonary resection," *J. Thoracic Cardiovascular Surg.*, vol. 159, no. 3, pp. 1130–1138, Mar. 2020.

[3] H. Teske, P. Mercea, M. Schwarz, N. H. Nicolay, F. Sterzing, and R. Bendl, "Real-time markerless lung tumor tracking in fluoroscopic video: Handling overlapping of projected structures," *Med. Phys.*, vol. 42, no. 5, pp. 2459–2540, 2015.

[4] D. Zhou, M. Nakamura, N. Mukumoto, M. Yoshimura, and T. Mizowaki, "Development of a deep learning-based patient-specific target contour prediction model for markerless tumor positioning," *Med. Phys.*, vol. 49, no. 3, pp. 1382–1390, Mar. 2022.

[5] W. Takahashi, S. Oshikawa, and S. Mori, "Real-time markerless tumour tracking with patient-specific deep learning using a personalised data generation strategy: Proof of concept by phantom study," *Brit. J. Radiol.*, vol. 93, no. 1109, May 2020, Art. no. 20190420.

[6] P. Markelj, D. Tomaževic, B. Likar, and F. Pernuš, "A review of 3D/2D registration methods for image-guided interventions," *Med. Image Anal.*, vol. 16, no. 3, pp. 642–661, Apr. 2012.

[7] C. J. F. Reyneke, M. Luthi, V. Burdin, T. S. Douglas, T. Vetter, and T. E. M. Mutsvangwa, "Review of 2-D/3-D reconstruction using statistical shape and intensity models and X-ray image synthesis: Toward a unified framework," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 269–286, 2019.

[8] J. Wang *et al.*, "Dynamic 2-D/3-D rigid registration framework using point-to-plane correspondence model," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1939–1954, Sep. 2017.

[9] H. Liao, W.-A. Lin, J. Zhang, J. Zhang, J. Luo, and S. K. Zhou, "Multiview 2D/3D rigid registration via a point-of-interest network for tracking and triangulation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12630–12639.

[10] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[11] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vis.*, vol. 61, no. 2, pp. 139–157, 2005.

[12] J. Rühaak *et al.*, "Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration," *IEEE Trans. Med. Imag.*, vol. 36, no. 8, pp. 1746–1757, Aug. 2017.

[13] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019.

[14] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Med. Image. Anal.*, vol. 52, pp. 128–143, Feb. 2019.

[15] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi, "Learning a probabilistic model for diffeomorphic registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2165–2176, Sep. 2019.

[16] S. Zhao, T. Lau, J. Luo, I. Eric, C. Chang, and Y. Xu, "Unsupervised 3D end-to-end medical image registration with volume tweening network," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 5, pp. 1394–1404, May 2020.

[17] K. Tang, Z. Li, L. Tian, L. Wang, and Y. Zhu, "ADMIR–affine and deformable medical image registration for drug-addicted brain images," *IEEE Access*, vol. 8, pp. 70960–70968, 2020.

[18] Y. Lei *et al.*, "4D-CT deformable image registration using multiscale unsupervised deep learning," *Phys. Med. Biol.*, vol. 65, no. 8, Apr. 2020, Art. no. 085003.

[19] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2CT-GAN: Reconstructing CT from biplanar X-rays with generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10611–10620.

[20] Y. Kasten, D. Doktofsky, and I. Kovler, "End-to-end convolutional neural network for 3D reconstruction of knee bones from bi-planar X-ray images," in *Proc. Int. Workshop Mach. Learn. Med. Image Reconstruction*, Lima, Peru, 2020, pp. 123–133.

[21] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2463–2471.

[22] N. Wang *et al.*, "Pixel2Mesh: 3D mesh model generation via image guided deformation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3600–3613, Oct. 2021.

[23] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," Presented at the 5th Int. Conf. Learn. Represent., 2017.

[24] K. Lin, L. Wang, and Z. Liu, "End-to-end human pose and mesh reconstruction with transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1954–1963.

[25] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, May 2016.

[26] S. Miao *et al.*, "Dilated FCN for multi-agent 2D/3D medical image registration," in *Proc. AAAI* 2018, pp. 1–8.

[27] R. Schaffert, J. Wang, P. Fischer, A. Borsdorf, and A. Maier, "Learning an attention model for robust 2-D/3-D registration using point-to-plane correspondences," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3159–3174, Oct. 2020.

[28] M. Nakao, Y. Oda, K. Taura, and K. Minato, "Direct volume manipulation for visualizing intraoperative liver resection process," *Comput. Methods Programs Biomed.*, vol. 113, no. 3, pp. 725–735, Mar. 2014.

[29] A. Saito, M. Nakao, Y. Uranishi, and T. Matsuda, "[POSTER] deformation estimation of elastic bodies using multiple silhouette images for endoscopic image augmentation," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Sep. 2015, pp. 170–171.

[30] B. Koo, E. Özgur, B. L. Roy, E. Buc, and A. Bartoli, "Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2017, pp. 326–334.

[31] M. D. Ketcha *et al.*, "Multi-stage 3D–2D registration for correction of anatomical deformation in image-guided spine surgery," *Phys. Med. Biol.*, vol. 62, no. 11, pp. 4604–4622, Jun. 2017.

[32] R. Modrzejewski, T. Collins, A. Bartoli, A. Hostettler, and J. Marescaux, "Soft-body registration of pre-operative 3D models to intra-operative RGBD partial body scans," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2018, pp. 39–46.

[33] M. Nakao, M. Nakamura, and T. Matsuda, "Image-to-graph convolutional network for deformable shape reconstruction from a single projection image," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2021, pp. 259–268.

[34] K. M. Brock, J. M. Balter, L. A. Dawson, M. L. Kessler, and C. R. Meyer, "Automated generation of a four-dimensional model of the liver using warping and mutual information," *Med. Phys.*, vol. 30, no. 6, pp. 1128–1133, 2003.

[35] C.-R. Chou and S. Pizer, "Real-time 2D/3D deformable registration using metric learning," in *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging* (Lecture Notes in Computer Science), vol. 7766, 2013, pp. 1–10.

[36] U. Mitrović, Ž. Špiclin, B. Likar, and F. Pernuš, "3D-2D registration of cerebral angiograms: A method and evaluation on clinical images," *IEEE Trans. Med. Imag.*, vol. 32, no. 8, pp. 1550–1563, Aug. 2013.

[37] T. De Silva *et al.*, "3D–2D image registration for target localization in spine surgery: Investigation of similarity metrics providing robustness to content mismatch," *Phys. Med. Biol.*, vol. 61, no. 8, pp. 3009–3025, Apr. 2016.

[38] U. Mitrović, B. Likar, F. Pernuš, and Ž. Špiclin, "3D–2D registration in endovascular image-guided surgery: Evaluation of state-of-the-art methods on cerebral angiograms," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 2, pp. 193–202, Feb. 2018.

[39] A. Lange and S. Heldmann, "Multilevel 2D–3D intensity-based image registration," in *Proc. Int. Workshop Biomed. Image Registration*, 2020, pp. 57–66.

[40] M. Nakao and K. Minato, "Physics-based interactive volume manipulation for sharing surgical process," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 809–816, May 2010.

[41] S. Suwelack *et al.*, "Physics-based shape matching for intraoperative image guidance," *Med. Phys.*, vol. 41, no. 11, Oct. 2014, Art. no. 111901.

[42] J. Ehrhardt, R. Werner, A. Schmidt-Richberg, and H. Handels, "Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 251–265, Feb. 2011.

[43] M. Nakao, J. Tokuno, T. Chen-Yoshikawa, H. Date, and T. Matsuda, "Surface deformation analysis of collapsed lungs using model-based shape matching," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 10, pp. 1763–1774, Oct. 2019.

[44] M. Nakamura *et al.*, "Statistical shape model-based planning organ-at-risk volume: Application to pancreatic cancer patients," *Phys. Med. Biol.*, vol. 66, no. 1, Jan. 2021, Art. no. 014001.

[45] M. Nakao, M. Nakamura, T. Mizowaki, and T. Matsuda, "Statistical deformation reconstruction using multi-organ shape features for pancreatic cancer localization," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101829.

[46] D. Toth *et al.*, "3D/2D model-to-image registration by imitation learning for cardiac procedures," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, no. 8, pp. 1141–1149, Aug. 2018.

[47] S. Wu, M. Nakao, J. Tokuno, T. Chen-Yoshikawa, and T. Matsuda, "Reconstructing 3D lung shape from a single 2D image during the deaeration deformation process using model-based data augmentation," *IEEE Int. Conf. Biomed. Health Inform. (BHI)*, May 2019, pp. 1–4.

[48] Y. Wang, Z. Zhong, and J. Hua, "Deeporgannet: On-the-fly reconstruction and visualization of 3D/4D lung models from single-view projections by deep deformation network," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 960–970, Jan. 2020.

[49] F. Tong, M. Nakao, S. Wu, M. Nakamura, and T. Matsuda, "X-ray2Shape: Reconstruction of 3D liver shape from a single 2D projection image," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 1608–1611.

[50] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[51] M. Akimoto *et al.*, "Inter- and intrafractional variation in the 3-dimensional positions of pancreatic tumors due to respiration under real-time monitoring," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 98, no. 5, pp. 1204–1211, Aug. 2017.

[52] K. K. Brock, S. Mutic, T. R. McNutt, H. Li, and M. L. Kessler, "Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM radiation therapy committee task group, no. 132," *Med. Phys.*, vol. 44, no. 7, pp. e43–e76, Jul. 2017.

[53] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.