Multi-Atlas Image Soft-Segmentation via Computation of the Expected Label Value

Iman Aganj and Bruce Fischl

***Abstract*—The use of multiple atlases is common in medical image segmentation. This typically requires deformable registration of the atlases (or the average atlas) to the new image, which is computationally expensive and susceptible to entrapment in local optima. We propose to instead consider the probability of all possible atlas-to-image transformations and compute the *expected label value (ELV)*, thereby not relying merely on the transformation deemed "optimal" by the registration method. Moreover, we do so without actually performing deformable registration, thus avoiding the associated computational costs. We evaluate our ELV computation approach by applying it to brain, liver, and pancreas segmentation on datasets of magnetic resonance and computed tomography images.**

***Index Terms*—Expected label value (ELV), supervised image segmentation, soft segmentation, atlas, MRI, CT.**

I. Introduction

AUTOMATIC image segmentation is often a central step in medical imaging studies, enabling the analysis of specific regions of interest (ROIs). In supervised segmentation, an algorithm segments a new image using the information derived from a training dataset of images that are accompanied with ground-truth (e.g. manually delineated) ROI labels. Two popular approaches to supervised image segmentation use multiple atlases [1-3] and deep neural networks [4, 5]. In multi-atlas-based segmentation of a new image, atlas images are (or a mean template image is) deformably registered to the new (to-be-segmented) image. The manual labels are then propagated into the new image space using the computed transformations, and fused to form the new labels.

Deformable registration of the atlas images to the new image is computationally very demanding (except for recent deep-learning based approaches [6-9]) and is the bottleneck of atlas-based segmentation. To improve computational efficiency, it has been proposed to use only a subset of atlases [10], albeit at the price of discarding a portion of the available training data.

The transformation resulting from registration guides label propagation from the atlas to the new image. Being an iterative non-convex optimization, image registration is prone to becoming trapped in local optima, potentially leading to inaccurate propagation of the labels. Moreover, different but equally reasonable transformations may produce similar values for the registration objective function (within its margin of error). Thus, even if the global optimum is found, choosing it as the single correct transformation would disregard valuable information provided by other potentially valid transformations. Such a globally optimal solution is also rarely robust, as it is sensitive to disturbances of or changes to input images, or variations in acquisition parameters. To alleviate this issue, uncertainty in registration has been incorporated into Bayesian segmentation by approximating the marginalization over registration parameters via Markov Chain Monte Carlo techniques [11], which, even though efficiently implemented, would further increase the computational costs. Local measures of uncertainty in deformable registration have also been used to improve the sensitivity of the label propagation in atlas-based segmentation [12, 13].

In this work, we present a new atlas-based image soft-segmentation method that – instead of attempting to determine a single correct label – produces the expected value of the label at each voxel of the new image, while considering the probability of possible atlas-to-image transformations. This is accomplished without either explicitly sampling from the transformation distribution (which would be intractable) or running the costly deformable registration in training or testing stages. We create a single image from the training data, which we call the *key*. Then, for a new image (after affine alignment, if necessary), we compute the *expected label value (ELV)* map simply via a convolution with the key, which is efficiently performed using the fast Fourier transform (FFT). Our fuzzy ELV map is therefore a robust combination of labels suggested

Support for this research was provided by the National Institutes of Health (NIH), specifically the National Institute of Diabetes and Digestive and Kidney Diseases (K01DK101631), the National Institute on Aging (R56AG068261, R01AG022381, R01AG016495), the National Institute for Biomedical Imaging and Bioengineering (P41EB015896, R01EB019956, R01EB023281), and the National Institute for Neurological Disorders and Stroke (R01NS105820, R01NS083534, U01NS086625). Additional support was provided by the BrightFocus Foundation (A2016172S).

Computational resources were provided by NIH Shared Instrumentation Grants (S10RR023401, S10RR019307, S10RR023043, S10RR028832), the O2 High Performance Compute Cluster at Harvard Medical School, the Enterprise Research Infrastructure & Services at Mass General Brigham (MGB), and the AWS Cloud Credits for Research program.

I. Aganj (iaganj@mgh.harvard.edu, +1-617-724-5652) and B. Fischl (bfischl@mgh.harvard.edu) are with the Athinoula A. Martinos Center for Biomedical Imaging, Radiology Department, Massachusetts General Hospital (MGH), Harvard Medical School, Charlestown, MA 02129, USA, and also with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

B. Fischl has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. B. Fischl's interests were reviewed and are managed by MGH and MGB according to their conflict of interest policies.

by atlas-to-image transformations, weighted by a measure of the transformation validity. This soft segmentation can be further used to initiate a subsequent hard-segmentation procedure. We validate our approach through brain, liver, and pancreas segmentation experiments on magnetic resonance (MR) and computed tomography (CT) images.

This article extends our preliminary conference version [14]. In particular, we have improved the method as well as expanded our empirical evaluation by including several new datasets. Moreover, our Matlab toolbox is now publicly available (https://www.nitrc.org/projects/elv). In the following, we describe the proposed method in detail (Section II and the appendices) and present experimental results (Section III) along with some concluding remarks (Section IV).

II. Methods

A. Segmentation from a Single Atlas

Let $I\colon \mathbb{R}^d \to \mathbb{R}$ be the d-dimensional image to be segmented, and $J\colon \mathbb{R}^d \to \mathbb{R}$ an atlas image with the same contrast as I, for which the manual label of a specific ROI has been provided as $L\colon \mathbb{R}^d \to \{0,1\}$.[1] For the new image I, we wish to compute the expected value of the ROI label, $E\colon \mathbb{R}^d \to [0,1]$, which is a measure of likelihood of each voxel belonging to the ROI.

In traditional atlas-based image segmentation, the label L is propagated into the space of I as $L \circ \boldsymbol{T}^*(I,J)$, where the transformation $\boldsymbol{T}^*(I,J)$ is computed via registration as $\boldsymbol{T}\colon \mathbb{R}^d \to \mathbb{R}^d$ that maximizes the similarity between I and $J \circ \boldsymbol{T}$.[2] Here, instead, we propose to compute the *expected value* of the propagated L, while considering a probability for each possible transformation in $\mathbb{T} \coloneqq \{\boldsymbol{T}\colon \mathbb{R}^d \to \mathbb{R}^d\}$, as follows:

$$\begin{aligned} E &\coloneqq \mathbf{E}[L \circ \boldsymbol{T} | I,J] \\ &= \int_{\mathbb{T}} \Pr(\boldsymbol{T}|I,J)\,(L \circ \boldsymbol{T})\mathrm{d}\boldsymbol{T}. \end{aligned} \tag{1}$$

Equation (1) computes the ELV as an integral over the space of all transformations, which could be regarded as multiple (theoretically an infinite number of nested) integrals over the space of parameters representing $\boldsymbol{T}$. For free-form deformation, as considered here, Eq. (1) in fact includes a d-dimensional integral – with respect to the value of $\boldsymbol{T}(\boldsymbol{x})$ – for each $\boldsymbol{x} \in \mathbb{R}^d$. In standard atlas-based segmentation, $\Pr(\boldsymbol{T}|I,J)$ is considered a Dirac delta, $\delta\big(\boldsymbol{T} - \boldsymbol{T}^*(I,J)\big)$, whereas here we will consider a full probability distribution for it.

Using Bayes' theorem, we can write the probability of the transformation given both the new and atlas images as:

$$\Pr(\boldsymbol{T}|I,J) \propto \Pr(I,J|\boldsymbol{T})\,\Pr(\boldsymbol{T}), \tag{2}$$

where the two right-hand-side factors correspond to the image similarity and the transformation regularity, respectively. For the former, we opt to use the inner product of the image and the transformed atlas, since it is expected to be higher when the two images are well aligned:

$$\Pr(I,J|\boldsymbol{T}) \propto \int_{\mathbb{R}^d} I(\boldsymbol{x})(J \circ \boldsymbol{T})(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{3}$$

It is, however, well established that the inner product reflects the degree of alignment more effectively when only the *phase* information of the image is included [15, 16], which is how in practice we will proceed, as described in Section II.C. A discussion on our choice of the inner product of phase images as image similarity is provided in Appendix B. In the following, we first consider the case where $\boldsymbol{T}$ is only a translation.

1) Translation

For a translation, $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\Delta}$, the inner product in Eq. (3) becomes the cross-correlation of the image and the atlas, which is commonly used for image alignment [15, 16]:

$$\begin{aligned} \Pr(I,J|\boldsymbol{\Delta}) &\propto \int_{\mathbb{R}^d} I(\boldsymbol{x})J(\boldsymbol{x} - \boldsymbol{\Delta})\mathrm{d}\boldsymbol{x} \\ &= (I * \bar{J})(\boldsymbol{\Delta}), \end{aligned} \tag{4}$$

where $*$ denotes the convolution operator, and $\bar{J}(\boldsymbol{x}) \coloneqq J(-\boldsymbol{x})$. By assuming a flat prior for the shift (i.e., a constant $\Pr(\boldsymbol{\Delta})$) and combining Eqs. (1), (2), and (4), the ELV at voxel $\boldsymbol{y}$ will be:

$$E(\boldsymbol{y}) \propto \int_{\mathbb{R}^d} (I * \bar{J})(\boldsymbol{\Delta})L(\boldsymbol{y} - \boldsymbol{\Delta})\mathrm{d}\boldsymbol{\Delta}, \tag{5}$$

or,

$$\begin{aligned} E &\propto (I * \bar{J}) * L \\ &= I * (\bar{J} * L). \end{aligned} \tag{6}$$

In the second line, we exploited the associativity property of convolution, which leads to the following expression for the ELV:

$$E \propto I * A, \tag{7}$$

where we define and pre-compute the *key*, A, from the atlas, as:

$$A \coloneqq \bar{J} * L. \tag{8}$$

As can be seen, A is obtained by flipping the atlas image, blurring it by the label, and shifting it so the label ROI is roughly at the center.

Next, we will incorporate deformations in our transformation model.

2) Deformation

To generalize the transformation $\boldsymbol{T}$ to include deformations, we will use the common Tikhonov prior on the regularity of the

[1] The ground-truth segmentation may also be a soft label, $L\colon \mathbb{R}^d \to [0,1]$.
[2] We denote vector-valued variables in bold.

deformation field as the probability of the transformation:

$$\Pr(\boldsymbol{T}) \propto R(\boldsymbol{T}) \coloneqq e^{-\frac{1}{2\sigma^2}\int_{\mathbb{R}^d}\|\partial \boldsymbol{T}(\boldsymbol{z})-\mathbb{I}\|_F^2 \mathrm{d}\boldsymbol{z}}, \tag{9}$$

where $\partial\boldsymbol{T}$ is the Jacobian matrix of $\boldsymbol{T}$, $\mathbb{I}$ is the $d \times d$ identity matrix, and the constant parameter σ represents a prior on the magnitude of the deformations. In Appendix A, we show that the ELV is still computed following Eq. (7), where the key, A, is initially computed as in Eq. (8), but then updated to incorporate the deformation. We show that we can approximate this update by an inhomogeneous blurring of the key, as:

$$A(\boldsymbol{x}) \leftarrow [A(\boldsymbol{z}) * G(\boldsymbol{z}|\boldsymbol{0}, \|\boldsymbol{x}\|_2\sigma^2\mathbb{I})]_{\boldsymbol{z}=\boldsymbol{x}}, \tag{10}$$

where $G(\cdot|\boldsymbol{\mu},\Sigma)$ represents the Gaussian function with the mean $\boldsymbol{\mu}$ and the co-variance matrix Σ. One can see that the size of the blurring kernel increases with the square root of the Euclidean distance from the center of A – i.e., the region corresponding to the label ROI (see Section II.A.1). Blurring a region in A decreases its contribution to soft segmentation by removing its high-frequency components prior to the convolution in Eq. (7). This means that the proposed ELV takes local deformations into account by giving a smaller weighting to regions in the atlas image that are farther from the ROI, making the information in such far areas less important.

The proposed model accounts for large translations, as well as local deformations, even though we do not run any deformable registration. As for rotation and global scaling, accounting for local deformations covers a small amount of them, and to allow for large amounts, we can initially affinely align the image and the atlas.

B. Multiple Atlases

In case N atlases (affinely normalized in the same space) with manual labels are available, we will write Eq. (1) in the same fashion, as:

$$E \coloneqq \frac{1}{N}\sum_{i=1}^{N} \mathbf{E}[L_i \circ \boldsymbol{T}|I, J_i], \tag{11}$$

where J_i and L_i are the i^{th} pair of atlas and manual-label images, respectively. This will yield similar results as in Eqs. (7) and (10), with the only difference being Eq. (8), now generalized as:

$$A \coloneqq \frac{1}{N}\sum_{i=1}^{N} \bar{J}_i * L_i. \tag{12}$$

Note that even in the case of multiple atlases, A is a *single* image that is pre-computed from the training data.

C. Implementation

1) Computation in the Fourier Domain

To create the key, A, we first ensure that the N training images are represented roughly in the same space; and if not, we affinely align them. By applying the convolution theorem to Eq. (12), we will then use FFT to initialize A:

$$A = \mathcal{F}^{-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{J}_i^*}{|\hat{J}_i|}\hat{L}_i\right\}, \tag{13}$$

where the *hat* $(\hat{\ })$ sign and $\mathcal{F}^{-1}$ represent the Fourier and inverse Fourier transforms, respectively, and $\hat{J}^*$ is the complex conjugate of $\hat{J}$. By only keeping the phase information of the image (i.e., normalizing $\hat{J}_i$ by its magnitude), we create a sharper probability distribution for the aligning transformation in Eq. (3) [15, 16] (see Appendix B). In addition, this has an intensity normalization effect, preventing A from giving a different weighting to an atlas image due to its global intensity scaling. Next, to incorporate deformations (i.e., if $\sigma > 0$), we update the key, A, voxel-wise following Eq. (10) by multiplying and summing it with a varying discretized Gaussian kernel.

To segment a new image, I, we first make sure that it is correctly represented in the atlas space (otherwise, affinely align it to the mean atlas image), and then compute the ELV map from Eq. (7) as follows:

$$E \propto \mathcal{F}^{-1}\left\{\frac{\hat{I}}{|\hat{I}|}\hat{A}\right\}. \tag{14}$$

Note that $\hat{A}$ is pre-computed from the atlases and kept offline.

For hard segmentation of the organ (or structure) from the map, we threshold the map to keep a voxel subset with the volume 14% larger than that of an average organ (estimated from the atlases); see Appendix C for the rationale behind this choice.[3] We then refine the mask by keeping the largest connected component (CC), as well as the CCs with at least half the volume of the largest CC, and then filling the holes.[4]

2) Second Pass

Once the initial ELV map is obtained, it can be refined by recalculating Eq. (14) while this time prioritizing the initial soft-segmented area. In our experiments, for instance, we used weighted versions of $A(\boldsymbol{x})$ and $I(\boldsymbol{x})$, as $A(\boldsymbol{x})G(\boldsymbol{x}|\boldsymbol{0}, s^2\mathbb{I})$ and $I(\boldsymbol{x})[E(\boldsymbol{y}) * G(\boldsymbol{y}|\boldsymbol{0}, s^2\mathbb{I})]_{\boldsymbol{y}=\boldsymbol{x}}$, respectively, where the size of the Gaussian window ($2s$) was chosen to be roughly that of an average organ.

3) Intensity Prior

Given that using the phase image discards some image intensity information, one can further augment the computed ELV volume with image intensities. At a given voxel, the Bayes

[3] For data that is not too noisy, the organ size can also be estimated as the inflection point of the curve obtained by sorting the ELV map in descending order. Alternatively, the ELV map can be thresholded with a value optimized from the training data.

[4] A Markov random field prior on the voxel labels could also be used to encourage spatial regularity [3, 18].

formula implies:

$$\Pr(L|I,E) = \frac{\Pr(I|L,E)\Pr(L|E)}{\Pr(I|L,E)\Pr(L|E)+\Pr(I|\neg L,E)\Pr(\neg L|E)}, \qquad (15)$$

where L indicates that the given voxel belongs to the label, with $\neg$ the negation operator, and I and E are the values of the image intensity and the computed ELV at the voxel, respectively. Were it known whether the voxel is included in the label or not, the image intensity would be conditionally independent of the ELV; i.e. $\Pr(I|L,E) = \Pr(I|L)$ and $\Pr(I|\neg L,E) = \Pr(I|\neg L)$. Using the ELV for $\Pr(L|E)$ then leads to:

$$\Pr(L|I,E) = \frac{\Pr(I|L)\,E}{\Pr(I|L)\,E+\Pr(I|\neg L)\,(1-E)}, \qquad (16)$$

where $\Pr(I|L)$ and $\Pr(I|\neg L)$ can be approximated by Gaussian functions of the intensity values of I, with their parameters estimated from the atlases (or the image itself using the initial ELV map). For E to exhibit the properties of a probability, the ELV map needs to be normalized by its maximum, with any negative values projected to zero.

$\Pr(L|I,E)$ can even substitute for I itself in the computation of the ELV, as – depending on the image contrast – it may better highlight the organ of interest, which is the most informative part of the image for segmentation. In that case, since the ELV map has not yet been computed, we use a constant E in Eq. (16) equal to the label-to-image volume ratio estimated from the atlases.

Several other post-processing steps are possible after this soft-segmentation [1]. If binary segmentation is desired, the ELV map can then be thresholded (see Appendix C and Section II.C.1) or used as seed region to subsequently initialize an unsupervised hard segmentation algorithm [14, 17].

III. Results and Discussion

We evaluated our ELV computation method on several medical image databases via leave-one-out cross validation. For each test image in a database, we created the key from the remainder of the images (i.e., labeled atlases) in the database following Eq. (12), computed the label for the test image, and report the Dice overlap coefficient between the computed label and the known label. Given that the images in each database were correctly represented in the same space, we did not affinely register them. Furthermore, since optimizing for σ in Eq. (10) improved the Dice scores only negligibly (< 1%) in our initial benchmarking, we report our results in this section for the simple case with $\sigma = 0$.

As described in Section II.C, we computed the ELV map in two passes, modulated the ELV with the intensity prior in Eq. (15), where we used the atlases to estimate the means (except for the liver; see Section III.B) and standard deviations, and then hard-thresholded the probability maps to create masks.

Additional steps to preprocess the abdominal CT images included: smoothing the borders of each image, automatically removing the patient table (via thresholding the image and removing the lower-most one or two connected components), and using the intensity-prior image (Section II.C.3) instead of the abdominal image itself for ELV computation (thereby highlighting the organ amongst all other parts of the image).

TABLE I
Dice Coefficients between ELV and FreeSurfer Labels in Brain

Structure	Hemisphere	Median	Mean	SEM
Thalamus	Left	0.751	0.744	0.001
	Right	0.771	0.764	0.001
Caudate	Left	0.804	0.793	0.002
	Right	0.807	0.796	0.002
Putamen	Left	0.808	0.797	0.002
	Right	0.816	0.802	0.002
Pallidum	Left	0.758	0.685	0.005
	Right	0.769	0.743	0.003
Hippocampus	Left	0.793	0.784	0.001
	Right	0.806	0.799	0.001
Amygdala	Left	0.754	0.742	0.002
	Right	0.758	0.748	0.002
All	**Both**	**0.782**	**0.766**	**0.001**

A. Brain

We first assessed the ability of our method to imitate FreeSurfer [19] in segmenting brain subcortical structures. We used T1-weighted MR images of 1224 subjects from the third release in the *Open Access Series of Imaging Studies (OASIS-3)* [20], normalized to the size 256×256×256 with 1 mm³ isotropic resolution. We considered the FreeSurfer-generated labels for 12 subcortical structures (left and right thalamus, caudate, putamen, pallidum, hippocampus, and amygdala) as "silver" standard and tried to reproduce the segmentation for each image via the proposed ELV approach. The median, mean, and standard error of the mean (SEM) of the resulting cross-validation Dice scores between the labels generated by ELV and FreeSurfer are shown in Table I. Overall, the Dice score had a median of 0.782 and a mean of 0.766 ± (SEM) 0.001 across subjects and structures.

Since no manually delineated labels were used as the gold standard in this experiment, the results merely reveal how faithful the proposed approach is in reproducing FreeSurfer labels. For comparison, in a similar experiment [21], a U-Net type convolutional neural network (CNN) was trained on 581 FreeSurfer-segmented T1-weighted brain images. The authors' trained model produced mean Dice scores of 0.74 and 0.71 on two manually labeled test datasets. (The authors, however, did not compare the labels that they computed with FreeSurfer-generated labels.)

B. Liver

Next, we used the training dataset of the public *Liver Tumor Segmentation (LiTS) Challenge* [22], which includes contrast-enhanced abdominal CT images with manually delineated

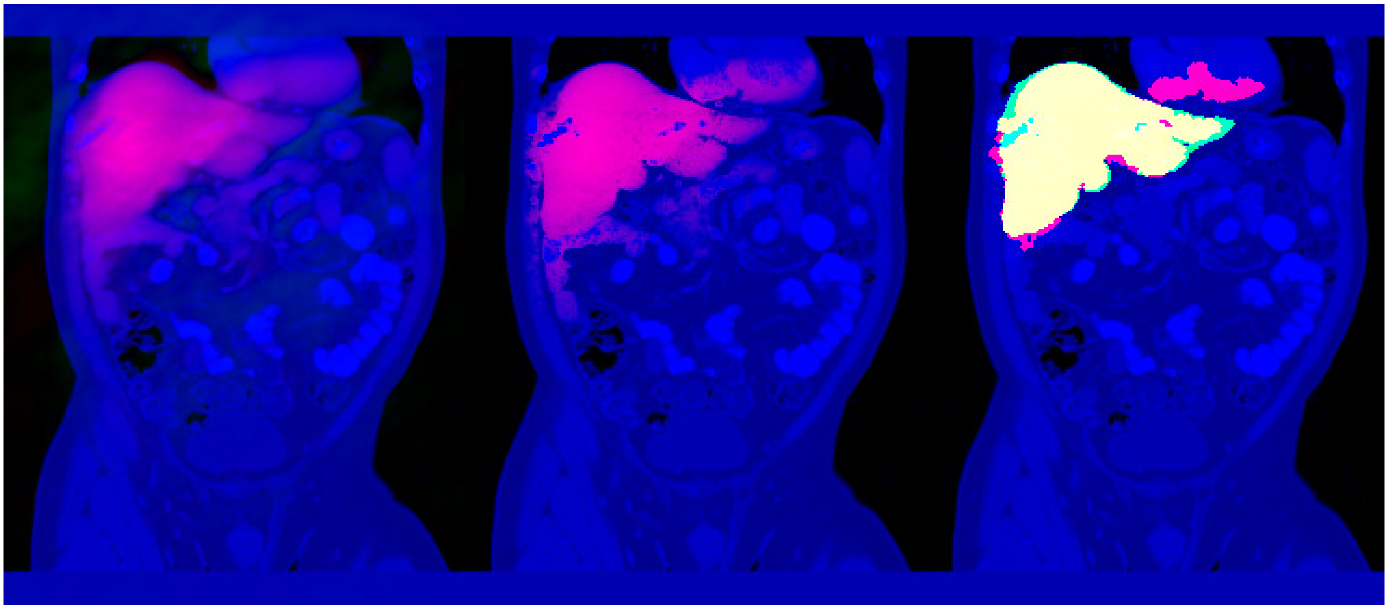

Fig. 1. CT image (blue) of the representative subject (i.e., with median segmentation Dice score) in the LiTS dataset. The slice with the largest cross section with the manual label is shown. *Left:* The ELV map of the liver (red; occasional negative values in green). *Middle:* The ELV map modulated by the intensity prior (red). *Right:* The resulting binary segmentation (red), the manual label (green), and their overlap (yellow). Intensities have been scaled for better visualization.

labels for the normal tissue and lesions in the liver, provided by various clinical sites. We considered the entire (healthy and lesion) organ label in our experiments. 85 subjects passed our inclusion criteria, mainly the slice thickness being included in the header and no larger than 2 mm. The images were resampled in the space of the first image to (1.6mm)³ isotropic resolution, so they were all of the size 248×248×323.

We then computed the ELV map for each subject, an example of which is illustrated in Fig. 1 (left) for the representative subject (corresponding to the median final Dice score; see below). To create the intensity-prior map, we estimated the standard deviation of the intensities of the liver and the background from the 84 atlas subjects, using the manual labels and their dilated versions (by a sphere of radius 50), respectively. For stability, we estimated the *mean* intensity using the initial ELV mask of the test subject, given that lesion size and intensity varied from subject to subject. Next, we modulated the ELV with the intensity prior (Fig. 1, middle), created a new mask, and further refined it with an updated intensity mean estimated from this mask. For mask preparation, we also performed morphological *opening* with a spherical structuring element with the radius of 2 voxels while keeping the largest connected component (i.e., eroding + keeping + dilating), which removed unwanted smaller structures attached to this relatively large ROI.

The cross-validation Dice coefficients between the computed masks and manual labels (Fig. 1, right) had a median of 0.92 across subjects (mean: 0.91 ± 0.01). A video of the ELV results for all subjects is available in the supplementary materials. Among those results with lower (entire-organ) Dice scores, lesion regions were frequently the culprit, as the intensity-prior map, although generally improving the segmentation, partially excluded some of those regions.

For comparison, we also trained a 3D CNN with the U-Net architecture [5] to segment the liver from 2-time downsampled LiTS images. The network consisted of 3 downsampling layers and 40 initial filters (at the first convolutional layer). We trained the network using 65536 3D sample patches of size 128×128×128 per epoch with a mini-batch size of 2. The CNN achieved a mean Dice score of 0.94 for the liver in cross-validation. Furthermore, at the time of the submission of this article, the LiTS challenge website [22] reported mean Dice values for the liver on their test data ranging from 0.84 to 0.97 (disregarding the outlier results with mean Dice ≤ 0.35), with many of the methods applying deep learning.

C. Pancreas

Lastly, we took a similar approach as in the previous subsection to segment the pancreas in two experiments, using two CT databases from *The Cancer Imaging Archive (TCIA)* acquired at the National Institutes of Health (NIH) Clinical Center [23, 24] (82 subjects) and from the *Memorial Sloan Kettering Cancer Center* [25] (225 subjects; those with slice thickness of 2~3 mm). The labels created from the ELV map in cross-validation and modulated with the intensity prior had a median Dice score of 0.59 (mean: 0.56 ± 0.02) for the former database and a median Dice score of 0.50 (mean: 0.48 ± 0.01) for the latter database. Note that the pancreas in the second dataset included lesions.

The pancreas' anatomical flexibility and variability in shape, size, and location make it a more challenging organ for segmentation than the liver and the brain subcortical structures, which could explain the lower accuracy of the results by our atlas-based method for this organ. For comparison, recent work using CNNs on the first (TCIA) dataset report Dice scores as high as 0.83 [24, 26, 27].

Note that, in contrast to mainstream supervised segmentation methods that employ deformable registration or sophisticated trained neural networks, we compute the ELV map via a simple linear convolution operation on the (phase) image.

IV. Conclusions

We have introduced a new approach to supervised soft-segmentation, which computes the expected label value (ELV) of a region of interest from an image using a training dataset of annotated atlases. The proposed method does not perform costly deformable registration, thereby also avoiding entrapment in local optima. We have evaluated the performance of our ELV computation technique in segmentation of the brain, the liver, and the pancreas. Future work consists of using the ELV map to augment the input to a convolutional neural network beyond the image itself, expecting to increase the segmentation accuracy of the better-informed model.

V. Appendices

A. Incorporation of Deformation

In this appendix, we derive the ELV while accounting for deformations in the transformation. By combining Eqs. (1), (2), (3), and (9), the ELV at voxel $\boldsymbol{y}$ will be:

$$E(\boldsymbol{y}) \propto \int_{\mathbb{R}^d} I(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \int_{\mathbb{T}} (J \circ \boldsymbol{T})(\boldsymbol{x})(L \circ \boldsymbol{T})(\boldsymbol{y})R(\boldsymbol{T})\mathrm{d}\boldsymbol{T}. \quad (17)$$

Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are fixed in the inner integral, we make the change of variables $\boldsymbol{T}(\boldsymbol{z}) = \boldsymbol{S}(\boldsymbol{z} - \boldsymbol{x})$. Note that such a global shift will not change either the regularization, i.e. $R(\boldsymbol{T}) = R(\boldsymbol{S})$, or the domain of the inner integral, $\mathbb{T}$. Consequently:

$$\begin{aligned} E(\boldsymbol{y}) &\propto \int_{\mathbb{R}^d} I(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \int_{\mathbb{T}} (J \circ \boldsymbol{S})(\boldsymbol{0})(L \circ \boldsymbol{S})(\boldsymbol{y} - \boldsymbol{x})R(\boldsymbol{S})\mathrm{d}\boldsymbol{S} \\ &= \int_{\mathbb{R}^d} I(\boldsymbol{x})A(\boldsymbol{y} - \boldsymbol{x})\mathrm{d}\boldsymbol{x}, \end{aligned} \quad (18)$$

or:

$$E \propto I * A, \quad (19)$$

where we define the *key*, A, as:

$$A(\boldsymbol{x}) := \int_{\mathbb{T}} (J \circ \boldsymbol{S})(\boldsymbol{0})(L \circ \boldsymbol{S})(\boldsymbol{x})R(\boldsymbol{S})\mathrm{d}\boldsymbol{S}. \quad (20)$$

Next, we write the transformation $\boldsymbol{S}$ as the sum of a global translation $\boldsymbol{\Delta} \in \mathbb{R}^d$ and a deformation (displacement) field $\boldsymbol{u} \in U$:

$$\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{\Delta} + \boldsymbol{u}(\boldsymbol{x}), \quad (21)$$

where $U := \left\{\boldsymbol{u}: \mathbb{R}^d \to \mathbb{R}^d \,\middle|\, \int_{\mathbb{R}^d} \boldsymbol{u}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \boldsymbol{0}\right\}$ is the set of translation-free displacement fields. The regularity prior is now:

$$R(\boldsymbol{S}) = \tilde{R}(\boldsymbol{u}) := e^{-\frac{1}{2\sigma^2}\int_{\mathbb{R}^d} \|\partial \boldsymbol{u}(\boldsymbol{z})\|_F^2 \mathrm{d}\boldsymbol{z}}. \quad (22)$$

We combine the above three equations, and separate the integral over the space of all transformations into an integral over possible translation-free deformations and an integral over possible translations:

$$A(\boldsymbol{x}) \propto \int_U \tilde{R}(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \int_{\mathbb{R}^d} J\big(\boldsymbol{\Delta} + \boldsymbol{u}(\boldsymbol{0})\big)L\big(\boldsymbol{x} + \boldsymbol{\Delta} + \boldsymbol{u}(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{\Delta}. \quad (23)$$

Note that this is a linear and invertible change of coordinates, hence $\mathrm{d}\boldsymbol{S} \propto \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{\Delta}$ (with the ratio independent of $\boldsymbol{S}$). With $\boldsymbol{u}$ and $\boldsymbol{x}$ being constant in the inner integral, we make the change of variables $\boldsymbol{\Delta} = \widetilde{\boldsymbol{\Delta}} - \boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{x}$, leading to:

$$\begin{aligned} A(\boldsymbol{x}) &\propto \int_U \tilde{R}(\boldsymbol{u})\mathrm{d}\boldsymbol{u} \int_{\mathbb{R}^d} J\big(\widetilde{\boldsymbol{\Delta}} - \boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{0}) - \boldsymbol{x}\big)L\big(\widetilde{\boldsymbol{\Delta}}\big)\mathrm{d}\widetilde{\boldsymbol{\Delta}} \\ &= \int_U \tilde{R}(\boldsymbol{u})A_0(\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}(\boldsymbol{0}) + \boldsymbol{x})\mathrm{d}\boldsymbol{u}, \end{aligned} \quad (24)$$

where A_0 is the key for the translation-only case, introduced in Eq. (8):

$$A_0 := \bar{J} * L. \quad (25)$$

It can be verified that:

$$\lim_{\sigma \to 0} A \propto A_0. \quad (26)$$

We now analytically estimate the key, A, as a function of A_0 for $\sigma > 0$. Combining Eqs. (22) and (24) leads to:

$$A(\boldsymbol{x}) \propto \int_U A_0(\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}(\boldsymbol{0}) + \boldsymbol{x})e^{-\frac{1}{2\sigma^2}\int_{\mathbb{R}^d}\|\partial \boldsymbol{u}(\boldsymbol{z})\|_F^2 \mathrm{d}\boldsymbol{z}}\mathrm{d}\boldsymbol{u}. \quad (27)$$

For simplicity, let us for now assume that $\boldsymbol{x}$ lies on the positive half of the first Cartesian coordinate axis, i.e., $\boldsymbol{x} = x\boldsymbol{v}_1$, where $\boldsymbol{v}_1$ is the unit vector in the direction of the first axis, and $x \geq 0$. We also define the line segment $Q_x := \{t\boldsymbol{v}_1 | 0 \leq t \leq x\}$. Accordingly:

$$\boldsymbol{u}(\boldsymbol{x}) - \boldsymbol{u}(\boldsymbol{0}) = \int_{Q_x} \partial \boldsymbol{u}(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}' = \int_0^x \partial_1 \boldsymbol{u}(t\boldsymbol{v}_1)\mathrm{d}t, \quad (28)$$

where $\partial_1 \boldsymbol{u}$ is the partial derivative of $\boldsymbol{u}$ in the direction of $\boldsymbol{v}_1$. Therefore:

$$\begin{aligned} A(x\boldsymbol{v}_1) \propto \int_{\partial U} A_0 &\left(x\boldsymbol{v}_1 + \int_0^x \partial_1 \boldsymbol{u}(t\boldsymbol{v}_1)\mathrm{d}t\right) \\ &\times e^{-\frac{1}{2\sigma^2}\int_{\mathbb{R}^d}\|\partial \boldsymbol{u}(\boldsymbol{z})\|_F^2 \mathrm{d}\boldsymbol{z}}\mathrm{d}(\partial \boldsymbol{u}). \end{aligned} \quad (29)$$

Note that we made further simplifying approximation by integrating over the space of the Jacobian of the deformation,

∂U, instead of the space of the deformation, U, itself.[5]

In Eq. (29), the only values of $\partial\boldsymbol{u}$ on which A_0 depends are $\partial_1\boldsymbol{u}(\boldsymbol{z})$ for $\boldsymbol{z} \in Q_x$. Thus, we separate the integral into the product of three integrals, the first one being:

$$A(x\boldsymbol{v}_1) \propto \int_{\partial_1\{\boldsymbol{u}:Q_x\to\mathbb{R}^d\}} A_0\left(x\boldsymbol{v}_1 + \int_0^x \partial_1\boldsymbol{u}(t\boldsymbol{v}_1)\mathrm{d}t\right) \times e^{-\frac{1}{2\sigma^2}\int_0^x\|\partial_1\boldsymbol{u}(t\boldsymbol{v}_1)\|_2^2\mathrm{d}t}\mathrm{d}(\partial_1\boldsymbol{u}), \tag{30}$$

and the second and third integrals are:

$$\int_{\partial_1\{\boldsymbol{u}:\mathbb{R}^d\setminus Q_x\to\mathbb{R}^d\}} e^{-\frac{1}{2\sigma^2}\int_{\mathbb{R}^d\setminus Q_x}\|\partial_1\boldsymbol{u}(\boldsymbol{z})\|_2^2\mathrm{d}\boldsymbol{z}}\mathrm{d}(\partial_1\boldsymbol{u}) \times \int_{\partial_{2,\dots,d}U} e^{-\frac{1}{2\sigma^2}\int_{\mathbb{R}^d}\|\partial_{2,\dots,d}\boldsymbol{u}(\boldsymbol{z})\|_F^2\mathrm{d}\boldsymbol{z}}\mathrm{d}(\partial_{2,\dots,d}\boldsymbol{u}), \tag{31}$$

which are integrals of normal distributions and therefore constant, hence not included in the expression for $A(x\boldsymbol{v}_1)$ in Eq. (30).

Calculation of $A(x\boldsymbol{v}_1)$ can be made notationally easier by approximating the inner integrals in Eq. (30) as Riemann sums. We divide $[0, x]$ into n equal intervals ($n \to \infty$), with $\mathrm{d}t \approx x/n$, and define:

$$\boldsymbol{q}_k \coloneqq \frac{x}{n}\partial_1\boldsymbol{u}\left(\frac{k}{n}x\boldsymbol{v}_1\right). \tag{32}$$

The integral is now approximated as:

$$A(x\boldsymbol{v}_1) \propto \int_{\mathbb{R}^{nd}} A_0\left(x\boldsymbol{v}_1 + \sum_{k=1}^{n}\boldsymbol{q}_k\right) \times e^{-\frac{1}{2x\sigma^2/n}\sum_{k=1}^{n}\|\boldsymbol{q}_k\|_2^2}\mathrm{d}\boldsymbol{q}_1 \dots \mathrm{d}\boldsymbol{q}_n. \tag{33}$$

This is, in fact, n consecutive convolutions of A_0 with a d-dimensional Gaussian,

$$A(x\boldsymbol{v}_1) \propto \left[A_0(\boldsymbol{z}) * \overbrace{G\left(\boldsymbol{z}\middle|\boldsymbol{0}, \frac{x}{n}\sigma^2\mathbb{I}\right) * \dots * G\left(\boldsymbol{z}\middle|\boldsymbol{0}, \frac{x}{n}\sigma^2\mathbb{I}\right)}^{n}\right]_{\boldsymbol{z}=x\boldsymbol{v}_1}. \tag{34}$$

Given that convolution of n identical Gaussians results in a Gaussian with n times the variance, we have:

$$A(x\boldsymbol{v}_1) \propto [A_0(\boldsymbol{z}) * G(\boldsymbol{z}|\boldsymbol{0}, x\sigma^2\mathbb{I})]_{\boldsymbol{z}=x\boldsymbol{v}_1}. \tag{35}$$

We now exploit the rotational invariance of the Gaussian in Eq. (35) and that of the Frobenius norm of the Jacobian in Eq. (27), to generalize Eq. (35) for any $\boldsymbol{x} \in \mathbb{R}^d$:

$$A(\boldsymbol{x}) \propto [A_0(\boldsymbol{z}) * G(\boldsymbol{z}|\boldsymbol{0}, \|\boldsymbol{x}\|_2\sigma^2\mathbb{I})]_{\boldsymbol{z}=\boldsymbol{x}}. \tag{36}$$

Equation (36) is indeed the update presented in Eq. (10). Despite our use of the convolution notation in Eq. (36), A is not computed via an actual convolution, because the co-variance matrix of the Gaussian kernel varies depending on $\boldsymbol{x}$, where the result of the convolution is evaluated.

B. Inner Product as the Image Similarity Metric

The inner product of the new image I and the transformed atlas image $J \circ \boldsymbol{T}$, which we have proposed as the image similarity metric in Eq. (3), is closely related to the sum-of-squared-differences (SSD) cost function that is commonly used in image registration:

$$\begin{aligned} SSD &\coloneqq \int_{\mathbb{R}^d}\left(I(\boldsymbol{x}) - (J\circ\boldsymbol{T})(\boldsymbol{x})\right)^2\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{R}^d}\left(I^2(\boldsymbol{x}) + (J\circ\boldsymbol{T})^2(\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x} \\ &\quad - 2\int_{\mathbb{R}^d} I(\boldsymbol{x})(J\circ\boldsymbol{T})(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{37}$$

In order to establish an equivalence between maximizing our inner-product similarity function and minimizing SSD, it would seem necessary to include in Eq. (3) the term $-½\int_{\mathbb{R}^d}\left(I^2(\boldsymbol{x}) + (J\circ\boldsymbol{T})^2(\boldsymbol{x})\right)\mathrm{d}\boldsymbol{x}$, which is not necessarily constant with respect to $\boldsymbol{T}$ due to local volume changes in the transformation. The extra terms that such an addition would introduce in $\Pr(\boldsymbol{T}|I,J)$ of Eq. (2), however, can be seen to be independent of the global-translation component of $\boldsymbol{T}$. Then, since an integral with respect to $\boldsymbol{T}$ can be taken separately with respect to a global translation value and translation-free displacement fields, as in Eqs. (21)–(23), the extra terms in $E(\boldsymbol{y})$ of Eq. (1) (resulting from the new translation-independent terms in $\Pr(\boldsymbol{T}|I,J)$) would be constant (independent of $\boldsymbol{y}$), and therefore unnecessary in the computation of the ELV. Consequently, quantifying the similarity between two images as their inner product, as adopted here, corresponds to the common use of the SSD cost function in deformable image registration.

As mentioned in Section II.A, we use only the phase information of the images in Eq. (3), and measure the image similarity with the following inner product:

$$\Pr(I,J|\boldsymbol{T}) = \int_{\mathbb{R}^d}\frac{\hat{I}(\boldsymbol{\omega})\widehat{J\circ\boldsymbol{T}}(\boldsymbol{\omega})^*}{\left|\hat{I}(\boldsymbol{\omega})\widehat{J\circ\boldsymbol{T}}(\boldsymbol{\omega})^*\right|}\mathrm{d}\boldsymbol{\omega}. \tag{38}$$

[5] This change of variables (integrating with respect to $\partial\boldsymbol{u}$ instead of $\boldsymbol{u}$) is linear due to the linearity of the differential operator ∂, as well as invertible due to the translation-free constraint on $\boldsymbol{u}$. We continue with the relaxing assumption that $\partial\boldsymbol{u}$ has independent elements. Nevertheless, for $d \geq 2$, the variable set $\partial\boldsymbol{u}$ is redundant and has a larger dimension than $\boldsymbol{u}$ does, with elements that are interdependent given the linear relationship $\boldsymbol{\nabla}\times\partial\boldsymbol{u} = \boldsymbol{0}$. As a result, for an exact solution, the integral must be taken with respect to an independent *subset* of the elements of $\partial\boldsymbol{u}$ that includes the (independent) set $\partial_1\boldsymbol{u}(Q_x)$.

Using only the phase of the images, as in Eq. (38), is more suitable for the estimation of $\Pr(I,J|\boldsymbol{T})$, as it produces sharper probability distributions [15, 16]. To demonstrate this via an example, let us model the transformation as a simple translation, $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{\Delta}$. The inner product therefore becomes the cross-correlation of the phase images, similar to Eq. (4), with Eq. (38) exhibiting the anticipated normality property, $\int_{\mathbb{R}^d} \Pr(I,J|\boldsymbol{\Delta})\,\mathrm{d}\boldsymbol{\Delta} = 1$ (although $\Pr(I,J|\boldsymbol{\Delta})$ can occasionally become negative). Subsequently, in the simplistic case where J is a shifted version of I, i.e. $J(\boldsymbol{x}) = I(\boldsymbol{x} - \boldsymbol{\Delta}_0)$, Eq. (38) will lead to $\Pr(I,J|\boldsymbol{\Delta}) = \delta(\boldsymbol{\Delta} - \boldsymbol{\Delta}_0)$, which is the exact desired distribution here.

Lastly, the inner product is zero for non-overlapping I and $J \circ \boldsymbol{T}$, which is a crucial property for the image similarity metric to have in ELV computation.

C. Volume Threshold

To threshold the computed probability map of the organ, such as the ELV, we sort the values of the map and keep the top v^* voxels, where the optimal v^* needs to be determined. Assuming that the ground-truth label has v_g voxels, we define $l(\eta)$ as the value of the ground-truth label at the top v^{th} voxel, where $v \coloneqq \eta v_g$. An ideal probability map, whose top v_g voxels are the ground-truth label, is expected to produce the following boxcar function:

$$l_0(\eta) \coloneqq \begin{cases} 1 & , \quad \eta \in [0 \quad 1] \\ 0 & , \qquad \text{o.w.} \end{cases}. \tag{39}$$

In practice, however, the transition to zero at $\eta = 1$ is less sharp due to inaccurately classified voxels, which we approximate with the following inverted sigmoid function:

$$l_\gamma(\eta) \coloneqq \frac{1+\gamma}{1+\gamma\left(1+\frac{1}{\gamma}\right)^{(1+\gamma)\eta}}, \tag{40}$$

where γ is a nonnegative constant. Note that $\lim_{\gamma\to 0} l_\gamma$ is the l_0 defined in Eq. (39) for the ideal probability map, and that $\lim_{\gamma\to\infty} l_\gamma(\eta) = e^{-\eta}$. Furthermore, the normality of l_γ, i.e. $\int_0^\infty l_\gamma(\eta)\mathrm{d}\eta = 1$, guarantees the expected property of $\int_0^\infty l_\gamma(v/v_g)\mathrm{d}v = v_g$.

Keeping the top v voxels results in a mask that overlaps with the ground-truth label with the following Dice similarity coefficient:

$$DSC = \frac{2\int_0^v l_\gamma(v'/v_g)\mathrm{d}v'}{v+v_g} = \frac{2\int_0^\eta l_\gamma(\eta')\mathrm{d}\eta'}{1+\eta} = \frac{2\log\left(\frac{\gamma+1}{\gamma+\left(\frac{\gamma}{\gamma+1}\right)^{(1+\gamma)\eta}}\right)}{\log\left(1+\frac{1}{\gamma}\right)(1+\eta)}. \tag{41}$$

One can verify that, depending on the value of γ, the η_γ^* that maximizes the above Dice score ranges from $\eta_0^* = 1$ to $\eta_\infty^* = -W_{-1}(-e^{-2}) - 2 = 1.146$, where W_k is the branch k of the Lambert W function. Therefore, according to this model, the optimal number of top voxels of the probability map to keep (to maximize Dice) is $v^* = \eta_\gamma^* v_g$. Choosing the nominal value of $\gamma = 1$ results in $\eta_1^* = 1.141$, which led us to keep the top subset of voxels with a volume 14% larger than that of an average organ (Section II.C.1). Note that subsequent keeping of only the largest connected components in the resulting mask reduces the number of false-positive voxels, further increasing the Dice score.

References

[1] J. E. Iglesias, and M. R. Sabuncu, “Multi-atlas segmentation of biomedical images: A survey,” *Medical Image Analysis,* vol. 24, no. 1, pp. 205-219, 2015.

[2] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. Bach Cuadra, “A review of atlas-based segmentation for magnetic resonance brain images,” *Computer Methods and Programs in Biomedicine,* vol. 104, no. 3, pp. e158-e177, 2011/12/01/, 2011.

[3] J. E. Iglesias, M. R. Sabuncu, I. Aganj, P. Bhatt, C. Casillas, D. Salat, A. Boxer, B. Fischl, and K. Van Leemput, “An algorithm for optimal fusion of atlases with different labeling protocols,” *NeuroImage,* vol. 106, pp. 451-463, 2015.

[4] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis,* vol. 42, pp. 60-88, 2017.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* pp. 234-241.

[6] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Transactions on Medical Imaging,* vol. 38, no. 8, pp. 1788-1800, 2019.

[7] J. Krebs, H. Delingette, B. Mailhé, N. Ayache, and T. Mansi, “Learning a Probabilistic Model for Diffeomorphic Registration,” *IEEE Transactions on Medical Imaging,* vol. 38, no. 9, pp. 2165-2176, 2019.

[8] 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,” *Medical Image Analysis,* vol. 52, pp. 128-143, 2019/02/01/, 2019.

[9] M. A. Morales, D. Izquierdo-Garcia, I. Aganj, J. Kalpathy-Cramer, B. R. Rosen, and C. Catana, “Implementation and Validation of a Three-dimensional Cardiac Motion Estimation Network,” *Radiology: Artificial Intelligence,* vol. 1, no. 4, pp. e180080, 2019.

[10] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, “Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy,” *NeuroImage,* vol. 46, no. 3, pp. 726-738, 2009/07/01/, 2009.

[11] J. E. Iglesias, M. R. Sabuncu, and K. Van Leemput, “Improved inference in Bayesian segmentation using Monte Carlo sampling: Application to hippocampal subfield volumetry,” *Medical Image Analysis,* vol. 17, no. 7, pp. 766-778, 2013.

[12] I. J. Simpson, M. W. Woolrich, and J. A. Schnabel, “Probabilistic segmentation propagation from uncertainty in registration,” in Proceedings of Medical Image Understanding and Analysis, 2011.

[13] M. P. Heinrich, I. J. A. Simpson, B. W. Papież, S. M. Brady, and J. A. Schnabel, “Deformable image registration by combining uncertainty estimates from supervoxel belief propagation,” *Medical Image Analysis,* vol. 27, pp. 57-71, 2016/01/01/, 2016.

[14] I. Aganj, and B. Fischl, “Expected label value computation for atlas-based image segmentation,” in Proc. IEEE International Symposium on Biomedical Imaging, Venice, Italy, 2019, pp. 334–338.

[15] C. Kuglin, and D. Hines, "The phase correlation image alignment methed," in Proc. Int. Conference Cybernetics Society, 1975, pp. 163-165.

[16] J. J. Pearson, D. C. Hines, S. Golosman, and C. D. Kuglin, "Video-Rate Image Correlation Processor," in 21st Annual Technical Symposium, 1977, pp. 9.

[17] I. Aganj, M. G. Harisinghani, R. Weissleder, and B. Fischl, "Unsupervised medical image segmentation based on the local center of mass," *Scientific Reports,* vol. 8, pp. 13012, 2018/08/29, 2018.

[18] M. R. Sabuncu, B. T. T. Yeo, K. V. Leemput, B. Fischl, and P. Golland, "A Generative Model for Image Segmentation Based on Label Fusion," *IEEE Transactions on Medical Imaging,* vol. 29, no. 10, pp. 1714-1729, 2010.

[19] B. Fischl, "FreeSurfer," *NeuroImage,* vol. 62, no. 2, pp. 774-781, 2012.

[20] A. F. Fotenos, A. Snyder, L. Girton, J. Morris, and R. Buckner, "Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD," *Neurology,* vol. 64, no. 6, pp. 1032-1039, 2005.

[21] A. G. Roy, S. Conjeti, D. Sheet, A. Katouzian, N. Navab, and C. Wachinger, "Error Corrective Boosting for Learning Fully Convolutional Networks with Limited Data," *Medical Image Computing and Computer Assisted Intervention − MICCAI 2017.* pp. 231-239.

[22] "LiTS – Liver Tumor Segmentation Challenge; https://competitions.codalab.org/competitions/17094," 2017.

[23] H. R. Roth, A. Farag, E. B. Turkbey, L. Lu, J. Liu, and R. M. Summers, "Data From Pancreas-CT," The Cancer Imaging Archive, 2016.

[24] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Medical Image Analysis,* vol. 45, pp. 94-107, 2018/04/01/, 2018.

[25] "Pancreas Tumor Database of Memorial Sloan Kettering Cancer Center – Medical Segmentation Decathlon; http://medicaldecathlon.com," 2018.

[26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," in Medical Imaging with Deep Learning (MIDL), 2018.

[27] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, "A Fixed-Point Model for Pancreas Segmentation in Abdominal CT Scans," *Medical Image Computing and Computer Assisted Intervention − MICCAI 2017.* pp. 693-701.