Deformable Registration Through Learning of Context-Specific Metric Aggregation

Enzo Ferrante[1,2(✉)], Puneet K. Dokania[1,4], Rafael Marini[1,3], and Nikos Paragios[1,3]

[1] Center for Visual Computing, CentraleSupelec, INRIA,
Universite Paris-Saclay, Paris, France
[2] Biomedical Image Analysis (BioMedIA) Group,
Imperial College London, London, UK
`ferrante.enzo@gmail.com`
[3] TheraPanacea, Paris, France
[4] University of Oxford, Oxford, UK

Abstract. We propose a novel weakly supervised discriminative algorithm for learning context specific registration metrics as a linear combination of conventional similarity measures. Conventional metrics have been extensively used over the past two decades and therefore both their strengths and limitations are known. The challenge is to find the optimal relative weighting (or parameters) of different metrics forming the similarity measure of the registration algorithm. Hand-tuning these parameters would result in sub optimal solutions and quickly become infeasible as the number of metrics increases. Furthermore, such hand-crafted combination can only happen at global scale (entire volume) and therefore will not be able to account for the different tissue properties. We propose a learning algorithm for estimating these parameters locally, conditioned to the data semantic classes. The objective function of our formulation is a special case of non-convex function, difference of convex function, which we optimize using the concave convex procedure. As a proof of concept, we show the impact of our approach on three challenging datasets for different anatomical structures and modalities.

1 Introduction

Deformable image registration is a highly challenging problem frequently encountered in medical image analysis. It involves the definition of a similarity criterion (data term) that, once endowed with a deformation model and a smoothness constraint, determines the optimal transformation to align two given images. We adopt a popular graphical model framework [5] to cast deformable registration as a discrete inference problem. The definition of the data term is among the

---

E. Ferrante and P.K. Dokania—Equal contribution.

most critical components of the registration process. It refers to a function that measures the (dis)similarity between images such as mutual information (MI) or sum of absolute differences (SAD). Metric learning in the context of image registration [2,8,9,11,18] is an alternative that aims to determine the most efficient means of image comparison (similarity measure) from labeled visual correspondences. Our approach can be considered as a specific case of metric learning where the idea is to efficiently combine the well studied mono/multi-modal metrics depending on the local context. We aim to learn the relative weighting from a given training dataset using a learning framework conditioned on prior semantic knowledge. We propose a novel *weakly supervised* discriminative learning framework, based structured support vector machines (SSVM) [13,15] and its extension to latent models LSSVM [16], to learn the relative weights of context specific metric aggregations.

Metric Learning. Various metric learning methods have been proposed in the context of image registration. Lee et al. [8] introduced a multi-modal registration algorithm where the similarity measure is learned such that the target and the correctly deformed source image receive high similarity scores. The training data consisted of pre-aligned images and the learning is performed at the patch level with an assumption that the similarity measure decompose over the patches. [2,9] proposed the use of sensitive hashing to learn a multi-modal metric. Similar to [8], they adopted a patch-wise approach. The dataset consisted of pairs of perfectly aligned images and a collection of positive/negative pairs of patches. Another patch-based alternative was presented by [14] where the training set consisted of non-aligned images with manually annotated patch pairs (landmarks). More recently, approaches based on convolutional neural networks started to gain popularity. Zagoruyko et al. [18] discussed CNN architectures to learn patch based similarity measures. One of them was then adopted in [11] to perform image registration. These methods require ground truth data in the form of correspondences (patches, landmarks or dense deformation fields), which is extremely difficult to obtain in real clinical data. Instead, our method is only based on segmentation masks.

Metric Aggregation. In contrast to the above approaches, our method aggregates standard metrics using contextual information. [3] showed, in fact, that using a multichannel registration method where a set of features is globally considered instead of a single similarity measure, produced robust registration compared to using individual features. However, they did not discuss how these features can be weighted. Following this, [4] proposed to estimate different deformation fields from each feature independently, and then compose them into final diffeomorphic transformation. Such strategy produces multiple deformation models (equal to number of metrics) which might be locally inconsistent. Thus, their combination may not be anatomically meaningful. Our method is most similar to Tang et al. [12], which generates a vector weight map that determines, at each spatial location, the relative importance of each constituent of the overall metric. However, the proposed learning strategy still requires ground

truth data in the form of correspondences (pre-registered images) which is not necessary in our case.

Contribution. We tackle the scenario where the ground truth deformations are not known a priori. We consider these deformation fields as latent variables, and devise an algorithm within the LSSVM framework [16]. We model the latent variable imputation problem as the deformable registration problem with additional constraints. In the end, we incorporate the learned aggregated metrics in a context-specific registration framework, where different weights are used depending on the structures being registered.

2   The Deformable Registration Problem

Let us assume a source three dimensional ($3D$) image I, a source $3D$ segmentation mask S^I and a target $3D$ image J. The segmentation mask is formed by labels $s_k \in \mathcal{C}$, where $\mathcal{C}$ is the set of classes. We focus on the 3D to 3D deformable registration problem. Let us also adopt without loss of generality a graphical model [5] for the deformable registration problem. A deformation field is sparsely represented by a regular grid graph $G = (V, E)$, where V is the set of nodes and E is the set of edges. Each node $i \in V$ corresponds to a control point p_i. Each control point p_i is allowed to move in the $3D$ space, therefore, can be assigned a label $\boldsymbol{d_i}$ from the set of $3D$ displacement vectors $\mathcal{L}$. Notice that each $3D$ displacement vector is a tuple defined as $\boldsymbol{d_i} = \{dx_i, dy_i, dz_i\}$, where dx, dy, and dz are the displacements in the x, y, and z directions, respectively. The deformation (labeling of the graph G) denoted as $D \in \mathcal{L}^{|V|}$ is associated to a set of nodes V, where each node is assigned a displacement vector $\boldsymbol{d_i}$ from the set $\mathcal{L}$. The new control point obtained when the displacement $\boldsymbol{d_i}$ is applied to the original control point p_i is denoted as $\bar{p}_i$. Let us define a patch $\bar{\Omega}_i^I$ on the source image I centered at the displaced control point $\bar{p}_i$. Similarly, we define Ω_i^J as the patch on the target image J centered at the original control point p_i, and $\bar{\Omega}_i^{S^I}$ as the patch on the input segmentation mask centered at the displaced control point $\bar{p}_i$. Using the above notations, we define the unary feature vector corresponding to the ith node for a given displacement vector $\boldsymbol{d_i}$ as $\mathcal{U}_i(\boldsymbol{d_i}, I, J) = (u_1(\bar{\Omega}_i^I, \Omega_i^J), \cdots, u_n(\bar{\Omega}_i^I, \Omega_i^J)) \in \mathbb{R}^n$, where n is the number of metrics (or similarity measures) and $u_j(\bar{\Omega}_i^I, \Omega_i^J)$ is the unary feature corresponding to the jth metric on the patches $\bar{\Omega}_i^I$ and Ω_i^J. In case of single metric, we define $n = 1$. Therefore, given a weight matrix $W \in \mathbb{R}^{n \times |\mathcal{C}|}$, where $W(i, j)$ denote the weight of the ith metric corresponding to the class j, the unary potential of the ith node for a given displacement vector $\boldsymbol{d_i}$ is computed as:

$$\bar{\mathcal{U}}_i(\boldsymbol{d_i}, I, J, S^I; W) = \mathbf{w}(\bar{c})^\top \mathcal{U}_i(\boldsymbol{d_i}, I, J) \in \mathbb{R}. \tag{1}$$

where, $\mathbf{w}(\bar{c}) \in \mathbb{R}^n$ is the $\bar{c}^{th}$ column of the weight matrix W and $\bar{c}$ is the most dominant class in the patch on the source segmentation mask $\bar{\Omega}_i^{S^I}$ obtained as $\bar{c} = \mathrm{argmax}_{c \in \mathcal{C}} f(\bar{\Omega}_i^{S^I}, c)$, with $f(\bar{\Omega}_i^{S^I}, c)$ being the number of voxels of class c in the patch $\bar{\Omega}_i^{S^I}$. Other criterion could be used to find the dominant class.

The pairwise clique potential between the control points p_i and p_j is defined as $\mathcal{V}(\boldsymbol{d_i}, \boldsymbol{d_j})$, where $\mathcal{V}(.,.)$ is the L_1 norm between the two input arguments. Thus, the multi-class energy function is:

$$\mathcal{E}(I, J, S^I, D; W) = \sum_{i \in V} \bar{\mathcal{U}}_i(\boldsymbol{d_i}, I, J, S^I; W) + \sum_{(i,j) \in E} \mathcal{V}(\boldsymbol{d_i}, \boldsymbol{d_j}) \qquad (2)$$

Then, the optimal deformation is obtained as $\hat{D} = \operatorname{argmin}_{D \in \mathcal{L}^{|V|}} \mathcal{E}(I, J, S^I, D; W)$. This problem is NP-HARD in general. Similar to [5], we adopt a pyramidal approach to solve the problem efficiently. We use FastPD [7] for the inference at every level of the pyramid. Notice that the energy function (2) is defined over the nodes of the sparse graph G. Once we obtain the optimal deformation $\hat{D}$, we estimate the dense deformation field using a free form deformation (FFD) model [10] in order to warp the input image.

3   Learning the Parameters

Knowing the weight matrix W a priori is non-trivial and hand tuning it quickly becomes infeasible as the number of metrics and classes increases. We propose an algorithm to learn W conditioned on the semantic labels assuming that in the training phase semantic masks are available for the source and the target images. Instead of learning the complete weight matrix at once, we learn the weights (or parameters) for each class $c \in \mathcal{C}$ individually. Now onwards, the weight vector $\mathbf{w}_c$ denotes a particular column of the weight matrix W, representing the weights corresponding to the cth class.

Training Data. Consider a dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1, \cdots, N}$, where $\mathbf{x}_i = (I_i, J_i)$, I_i is the source image and J_i is the target. Similarly, $\mathbf{y}_i = (S_i^I, S_i^J)$, where S_i^I and S_i^J are the segmentation masks for the source and target images. The size of each segmentation mask is the same as that of the corresponding images. As stated earlier, the segmentation mask is formed by the elements (or voxels) $s_k \in \mathcal{C}$, where $\mathcal{C}$ is the set of classes.

Loss Function. The loss function $\Delta(S^I, S^J) \in \mathbb{R}_{\geq 0}$ evaluates the similarity between the segmentation masks S^I and S^J. Higher $\Delta(.,.)$ implies higher dissimilarity. We use a dice based loss function as this is our evaluation criteria:

$$\Delta(S^I, S^J) = 1 - DICE(S^I, S^J) = 1 - (2 \sum_{i \in V} \frac{|\phi(p_i^I) \cap \phi(p_i^J)|}{|\phi(p_i^I)| + |\phi(p_i^J)|}), \qquad (3)$$

where, $\phi(p_i^I)$ and $\phi(p_i^J)$ are the patches at the control point p_i on the segmentation masks S^I and S^J, respectively, and $|.|$ represents cardinality. This approximation makes the dice decomposable over the nodes of G enabling a very efficient training.

Joint Feature Map. Given $\mathbf{w}_c$ for the c-th class, the deformation D and input $\mathbf{x}$, the multi-class function (2) can be trivially converted into class-based energy

---

Algorithm 1. The CCCP Algorithm.

1: $\mathcal{D}$, $\mathbf{w}_0$, C, α, η, the tolerance ϵ.
2: $t = 0$, $\mathbf{w}_t = \mathbf{w}_0$.
3: **repeat**
4:     For a given $\mathbf{w}_t$, impute latent variables $\hat{D}_i$ for each sample by solving (9).
5:     Update parameters $\mathbf{w}_{t+1}$ by optimizing the convex optimization problem (10).
6:     $t = t + 1$
7: **until** The objective function of the problem (7) does not decrease more than ϵ.

---

function as:

$$\mathcal{E}_c(\mathbf{x}, D; \mathbf{w}) = \mathbf{w}_c^\top \sum_{i \in V} \mathcal{U}_i(\boldsymbol{d_i}, \mathbf{x}) + w_p \sum_{(i,j) \in E} \mathcal{V}(\boldsymbol{d_i}, \boldsymbol{d_j}), \tag{4}$$

where $w_p \in \mathbb{R}_{\geq 0}$ is the parameter for the pairwise term. The final parameter vector $\mathbf{w} \in \mathbb{R}^{n+1}$ is the concatenation of $\mathbf{w}_c$ and w_p. Thus, the function (4) can be written as:

$$\mathcal{E}_c(\mathbf{x}, D; \mathbf{w}) = \mathbf{w}^\top \Psi(\mathbf{x}, D), \tag{5}$$

where $\Psi(\mathbf{x}, D) \in \mathbb{R}^{n+1}$ is the joint feature map defined as:

$$\Psi(\mathbf{x}, D) = \begin{pmatrix} \sum_{i \in V} \mathcal{U}_i^1(\boldsymbol{d_i}, \mathbf{x}) \\ \vdots \\ \sum_{i \in V} \mathcal{U}_i^n(\boldsymbol{d_i}, \mathbf{x}) \\ \sum_{(i,j) \in E} \mathcal{V}(\boldsymbol{d_i}, \boldsymbol{d_j}) \end{pmatrix} \tag{6}$$

Notice that the energy function (4) does not depend on the source segmentation mask S^I. The only use of the source segmentation mask in the energy function (2) is to obtain the dominant class which in this case is not required. However, we will shortly see that the source segmentation mask S^I plays a crucial role in the learning algorithm.

Latent Variables. Ideally, the dataset $\mathcal{D}$ must contain the ground truth deformations D corresponding to the source image I in order to compute the energy term defined in the Eq. (4). Since annotating the dataset with the ground truth deformation is non-trivial, we use them as the latent variables in our algorithm.

The Objective Function. Given $\mathcal{D}$, we learn the parameter $\mathbf{w}$ such that minimizing the energy function (4) leads to a deformation field which when applied to the source segmentation mask gives minimum loss with respect to the target segmentation mask. We denote $g(S, D)$ as the deformed segmentation when the dense deformation field obtained from D is applied to the segmentation mask S. Similarly to the latent SSVM [16], we optimize a regularized upper bound on the loss:

$$\min_{\mathbf{w},\{\xi_i\}} \frac{1}{2}||\mathbf{w}||^2 + \alpha||\mathbf{w} - \mathbf{w}_0||^2 + \frac{C}{N}\sum_i \xi_i,$$

$$s.t. \min_{D,\Delta(g(S_i^I,D),S_i^J)=0} \mathbf{w}^\top \Psi(\mathbf{x}_i,D) \leq \mathbf{w}^\top \Psi(\mathbf{x}_i,\bar{D}) - \Delta(g(S_i^I,\bar{D}),S_i^J) + \xi_i,$$

$$\forall \bar{D}, w_p \geq 0, \xi_i \geq 0, \forall i. \tag{7}$$

where, $\bar{D} = \mathrm{argmin}_D \mathcal{E}(\mathbf{x}_i, D; \mathbf{w})$. The above objective function minimizes an upper bound on the given loss, called slack (ξ_i). The effect of the regularization term is controlled by the hyper-parameter C. The second term is the proximity term to ensure that the learned $\mathbf{w}$ is close to the initialization $\mathbf{w}_0$. The effect of the proximity term is controlled by the hyperparameter α. Intuitively, for a given input-output pair, the constraints of the above objective function enforce that the energy corresponding to the best possible deformation field, in terms of both energy and loss (in order to be semantically meaningful), must always be less than or equal to the energy corresponding to any other deformation field with a margin proportional to the loss and some non negative slack.

The Learning Algorithm. The objective function (7) turns out to be a special case of non-convex functions (difference of convex), thus can be locally optimized using the well known CCCP algorithm [17]. The CCCP algorithm consist of three steps – (1) upperbounding the concave part at a given $\mathbf{w}$, which leads to an affine function in $\mathbf{w}$; (2) optimizing the resultant convex function (sum of convex and affine functions is convex); (3) repeating the above steps until the objective can not be further decreased beyond a given tolerance of ϵ. The complete CCCP algorithm for the optimization of (7) is shown Algorithm 1. The first step of upperbounding the concave functions (Line 4) is the same as the latent imputation step, which we call the *segmentation consistent registration* problem. The second step is the optimization of the resultant convex problem (Line 5), which is the optimization of the SSVM for which we use the well known cutting plane algorithm [6]. In what follows, we discuss these steps in detail.

Segmentation Consistent Registration. This step involves generating the best possible ground truth deformation field (unknown a priori) at a given $\mathbf{w}$, known as the latent imputation step. Since we optimize the dice loss, we formulate this step as an inference problem with additional constraints to ensure that the imputed deformation warps the image minimizing the loss between the deformed source and the target. Mathematically, for a given parameter vector $\mathbf{w}$, the latent deformation is imputed by solving:

$$\hat{D}_i = \underset{D \in \mathcal{L}^{|V|}, \Delta(g(S_i^I,D),S_i^J)=0}{\mathrm{argmin}} \mathbf{w}^\top \Psi(\mathbf{x}_i,D). \tag{8}$$

We relax the above problem as it is difficult and may not have a unique solution:

$$\hat{D}_i = \underset{D \in \mathcal{L}^{|V|}}{\mathrm{argmin}} \left(\mathbf{w}^\top \Psi(\mathbf{x}_i,D) + \eta \Delta(g(S_i^I,D),S_i^J) \right), \tag{9}$$

where, $\eta \geq 0$ controls the relaxation trade-off. Since the loss function used is decomposable, the above problem can be optimized using FastPD inference for the deformable registration with trivial modifications on the unary potentials.

Parameters Update. Given the imputed latent variables, the resultant objective is:

$$\min_{\mathbf{w},\{\xi_i\}} \frac{1}{2}||\mathbf{w}||^2 + \alpha||\mathbf{w} - \mathbf{w}_0||^2 + \frac{C}{N}\sum_i \xi_i,$$

$$s.t. \mathbf{w}^\top \Psi(\mathbf{x}_i, \hat{D}_i) \leq \mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{D}) - \Delta(g(S_i^I, \bar{D}), S_i^J) + \xi_i, \forall \bar{D}, w_p, \xi_i \geq 0, \forall i. \quad (10)$$

where, $\hat{D}_i$ is the latent deformation field imputed by solving the problem (9). Intuitively, the above objective function tries to learn the parameters $\mathbf{w}$ such that the energy corresponding to the imputed deformation field is always less than the energy for any other deformation field with a margin proportional to the loss function. The above objective function has exponential number of constraints, one for each possible deformation field $\bar{D} \in \mathcal{L}^{|V|}$. In order to alleviate this problem we use cutting plane algorithm [6]. Briefly, for a given $\mathbf{w}$, each deformation field $\bar{D}$ gives a slack. Instead of minimizing all the slacks for a particular sample at once, we find the deformation field that leads to the maximum value of the slack and store this in a set known as the working set. This is known as *finding the most violated constraint*. Thus, instead of using exponentially many constraints, the algorithm uses the constraints stored in the working set and this process is repeated until no constraints can be further added. Rearranging the terms in the constraints of the objective function (10) and ignoring the constant term $\mathbf{w}^\top \Psi(\mathbf{x}_i, \hat{D}_i)$, the most violated constraint can be obtained by solving:

$$\bar{D}_i = \operatorname*{argmin}_{D \in \mathcal{L}^{|V|}} \left(\mathbf{w}^\top \Psi(\mathbf{x}_i, \bar{D}) - \Delta(g(S_i^I, \bar{D}), S_i^J)\right). \quad (11)$$

Since the loss is decomposable, this problem can be solved using FastPD inference for the deformable registration with trivial modifications on the unary terms.

Prediction. Once we obtain the learned parameters $\mathbf{w}_c$ for each class $c \in \mathcal{C}$ using the Algorithm 1, we form the matrix W where each column of the matrix represents the learned parameter for a specific class. This W is then used to solve the registration problem (Eq. 2) using the approximate inference discussed in Sect. 2.

4   Results and Discussion

As a proof of concept, we evaluate the effect of the aggregated metric on three different medical datasets – (1) RT Parotids, (2) RT Abdominal, and a downsampled version of (3) IBSR [1], involving several anatomical structures, different image modalities, and inter/intra patient images We used four different metrics: (1) sum of absolute differences (SAD), (2) mutual information (MI), (3) normalized cross correlation (NCC), and (4) discrete wavelet coefficients (DWT). The datasets consists of 8 CT (RT Parotids, head images of $56 \times 62 \times 53$ voxels), 5 CT (RT Abdominal, abdominal images of $90 \times 60 \times 80$ voxels) and 18 MRI images (a downsampled version of IBSR dataset, including brain images $64 \times 64 \times 64$
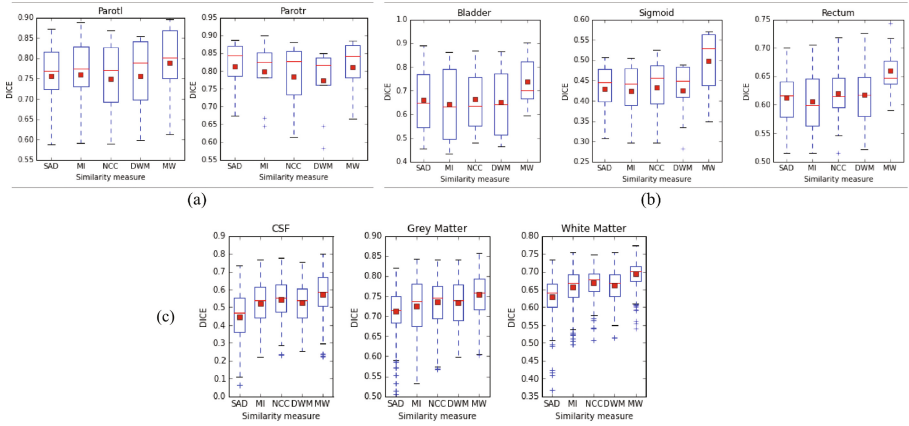
Fig. 1. Results for RT parotids (a), RT abdominal (b) and IBSR (c) datasets. We show dice between the deformed source and the target segmentation masks after registration, for the single-metric registration (SAD, MI, NCC, DWT) and the learned multi-metric registration (MW). In (a), 'Parotl' and 'Parotr' are the left and the right parotids. In (b), 'Bladder', 'Sigmoid', and 'Rectum' are the three organs in the dataset. In (c), annotations correspond to Cerebrospinal fluid (CSF), grey (GM) and white (WM) matter. The red square is the mean and the red bar is median. (Color figure online)

voxels). We performed multi-fold cross validation in every dataset, considering pairs with different patients in training and testing. For a complete description of the datasets and the experimental setting, please refer to the supplementary material. The results on the test sets are shown in Fig. 1. As it can be observed in Fig. 1, the linear combination of similarity measures weighted using the learned coefficients systematically outperforms (or is as good as) single metric based registration, with max improvements of 8% in terms of dice (see Fig. 2).
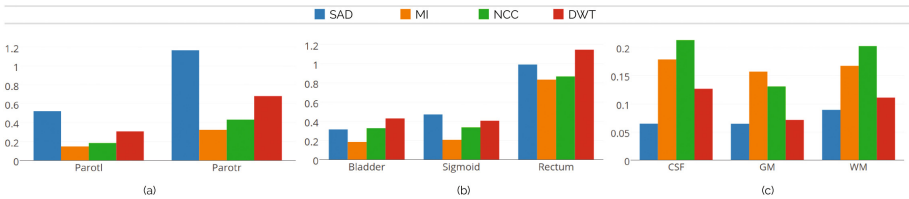
Fig. 2. Example of learned weights for RT Parotids (a), RT Abdominal (b) and IBSR (c) datasets. Since the structures of interest in every dataset present different intensity distributions, different metric aggregations are learned. Note that in case of the RT Parotids, given that both parotid glands present the same intensity distribution, similar weightings are learned for both structures, with SAD dominating the other similarity measures. However, in IBSR dataset, NCC dominates in case of CSF and WM, while MI receives the higher value for gray matter (GM).

Discussion and Conclusions. We have showed that associating different similarity criteria to every anatomical region yields results superior to the classic single metric approach. In order to learn this mapping where ground truth is generally given in the form of segmentation masks, we defined deformation fields as latent variables and proposed a LSSVM based framework. The main limitation of our method is the need of segmentation masks for the source images in testing time. However, different real scenarios like radiation therapy or atlas-based segmentation methods fulfill this condition. Note that, at prediction (testing) time, the segmentation mask is used to determine the metrics weights combination per control node (finding the dominant class). The segmentation labels are not used at testing time to guide the registration process which is purely image based. In our multi-metric registration approach, segmentation masks are only required (at testing time) for the source image and used to choose the best learned metric aggregation. The idea could be further extended to unlabeled data (as it concerns the source image at testing time) where the dominant label class per control node is the output of a classification/learning method. From a theoretical viewpoint, we showed how the three main components of LSSVM: (1) latent imputation (Eq. 9); (2) prediction (optimizing Eq. 2) and (3) finding most violated constraint (Eq. (11)), can be formulated as the exact same problem. The difference among these problems is the unary potentials used. This is extremely important given that further improvements in inference algorithms will directly increase the quality of the results. As future work, the integration of alternative accuracy measures, other than dice, such as the Hausdorff distance between surfaces or real geometric distances for anatomical landmarks could further enhance the performance of the method.

References

1. IBSR: Internet Brain Segmentation Repository. http://www.cma.mgh.harvard.edu/ibsr/
2. Bronstein, M.M., Bronstein, A.M., Michel, F., Paragios, N.: Data fusion through cross-modality metric learning using similarity-sensitive hashing. In: CVPR 2010, pp. 3594–3601. IEEE (2010)
3. Cifor, A., Risser, L., Chung, D., Anderson, E.M., Schnabel, J.A.: Hybrid feature-based Log-Demons registration for tumour tracking in 2-D liver ultrasound images. In: ISBI (2012)
4. Cifor, A., Risser, L., Chung, D., Anderson, E.M., Schnabel, J.A.: Hybrid feature-based diffeomorphic registration for tumor tracking in 2-D liver ultrasound images. IEEE TMI **32**, 1647–1656 (2013)
5. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med. Image Anal. **12**(6), 731–741 (2008)
6. Joachims, T., Finley, T., Yu, C.: Cutting-plane training of structural SVMs. Mach. Learn. **77**, 27–59 (2009)
7. Komodakis, N., Tziritas, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic MRFs. In: CVPR (2007)

8. Lee, D., Hofmann, M., Steinke, F., Altun, Y., Cahill, N.D., Scholkopf, B.: Learning similarity measure for multi-modal 3D image registration. In: CVPR 2009, IEEE Conference on Computer Vision and Pattern Recognition, pp. 186–193. IEEE (2009)

9. Michel, F., Bronstein, M., Bronstein, A., Paragios, N.: Boosted metric learning for 3D multi-modal deformable registration. In: ISBI (2011)

10. Rueckert, D., Sonoda, L.I., et al.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE TMI **18**, 712–721 (1999)

11. Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., Komodakis, N.: A deep metric for multimodal registration. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 10–18. Springer, Cham (2016). doi:10.1007/978-3-319-46726-9_2

12. Tang, L., Hero, A., Hamarneh, G.: Locally-adaptive similarity metric for deformable medical image registration. In: ISBI. IEEE (2012)

13. Taskar, B., Guestrin, C., Koller, D.: Max-margin Markov networks. In: NIPS (2003)

14. Toga, A.W.: Learning based coarse-to-fine image registration. In: CVPR (2008)

15. Tsochantaridis, I., Hofmann, T., Joachims, T., Altun, Y.: Support vector machine learning for interdependent and structured output spaces. In: ICML (2004)

16. Yu, C.N., Joachims, T.: Learning structural SVMs with latent variables. In: ICML (2009)

17. Yuille, A., Rangarajan, A.: The concave-convex procedure. Neural Comput. **15**, 915–936 (2003)

18. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via convolutional neural networks