An Efficient Optimization Framework for Multi-Region Segmentation based on Lagrangian Duality

Johannes Ulén, Petter Strandmark and Fredrik Kahl

Abstract—We introduce a multi-region model for simultaneous segmentation of medical images. In contrast to many other models, geometric constraints such as inclusion and exclusion between the regions are enforced, which makes it possible to correctly segment different regions even if the intensity distributions are identical. We efficiently optimize the model using a combination of graph cuts and Lagrangian duality which is faster and more memory efficient than current state of the art. As the method is based on global optimization techniques, the resulting segmentations are independent of initialization. We apply our framework to the segmentation of the left and right ventricles, myocardium and the left ventricular papillary muscles in MRI and to lung segmentation in full-body X-ray CT. We evaluate our approach on a publicly available benchmark with competitive results.

Index Terms—Image segmentation, discrete optimization, cardiac segmentation, lung segmentation.

I. INTRODUCTION

The field of medical imaging is full of challenging segmentation tasks. The aim of the present paper is to segment multiple regions with a model that encompasses both the underlying appearance and shape of the different regions as well as their geometric relationships. This is often overlooked in present methods. For example, many successful approaches to cardiac segmentation concentrate on segmenting the left ventricle (LV) as this part is the most interesting for diagnostic purposes. Still, quantifiable information about the cardiac function is gained from segmenting the right ventricle (RV) as well. Our framework allows for the construction of a joint model of the whole heart where the final result is improved compared to segmenting the parts independently.

A. Optimizability and fidelity

Models in medical image analysis have two important desiderata: *optimizability* and *fidelity* [31]. Optimizability specifies how amenable the model is to optimization techniques. For instance, models involving curvature and geometric shape priors tend to be harder to optimize than models only based on image intensities. On the other hand, these more complex models often describe the reality better and have the potential for better results — they have higher fidelity. There is a natural trade-off between optimizability and fidelity. In this section, we categorize and review related work based on these two criteria.

Most segmentation approaches rely on local descent techniques, e.g. [28], [32], [35], and may get stuck in local optima

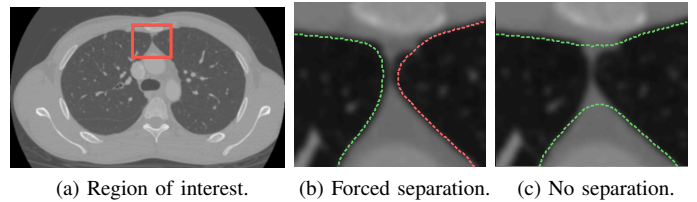(a) Region of interest.   (b) Forced separation.   (c) No separation.

Figure 1. In (a) the region of interest is marked with a rectangle. In (b) the model has constraints that force the two regions to be separate. In (c) the exclusion constraint is removed. This results in a segmentation where one region wrongly overflows into the other. Note that the image data for the correct segmentation is very weak and hence it is necessary to encode this prior information into the model.

which make them less reliable. These methods do not have good optimizability, but their fidelity can still be excellent. On the other hand, it is possible to apply global optimization techniques to make the segmentation more robust to poor initialization, e.g. [8], [27], [2]. These methods have great optimizability, but often the model is less sophisticated and impoverished. Another example is the random walker model [16]. Although the model is quite restricted, for example, there is no way to control the smoothness of the segmentation boundaries, the approach has proven effective for interactive segmentation.

During the last decade, optimization techniques for segmentation have greatly improved [11],[25],[26].In particular, graph cuts methods have become a standard tool, see the recent book [4] and the references therein. This is mainly due to their ability to compute globally optimal solutions in an efficient manner. However, the models are required to be submodular which is a strong limitation in practice and makes the models less faithful. More specifically, optimization problems for multi-region models are typically not submodular. An exception is the Markov random field model with convex priors introduced in [22]. In order to cope with more faithful multi-region models, so-called move-making algorithms have been developed that iteratively refines an initial segmentation where a submodular graph-cut problem is solved in each step [11]. The downside of such approaches is of course the dependence on a good initial solution.

The purpose of this paper is to increase the fidelity of models in medical image segmentation without sacrificing optimizability. For example, the human heart is composed of several interacting geometric parts — this fact should be reflected in the model. Secondly, the model should be complete in the sense that every voxel of the image should be

modeled, both in terms of geometry/shape and appearance using statistical principles. We are able to achieve this for cardiac segmentation in MRI while still computing close to optimal segmentations in an efficient manner. As another demonstration of the generality of the optimization framework, we consider the problem of segmenting lungs in full-body X-ray CT. Popular methods for this problem use multiple steps where first an initial segmentation is found, then the left and right lungs are separated in a subsequent step. In some cases [20], [3], the lungs are separated in each slice individually and in other cases they are separated for the whole volume at the same time [39]. In contrast, we are able to solve the resulting optimization problem for our model optimally, without having to resort to multiple phases. This is what we mean by increased optimizability without sacrificing fidelity, see Figure 1.

B. Contributions

The main contribution of our work is a multi-region segmentation framework with good optimizability. Our framework builds on the multi-region scheme presented in [14] where it is shown that geometric relationships, for example, when one object is included in another, can be modeled and globally optimized via graph cuts. The key property that makes this possible is that the resulting energy minimization problem is submodular. We also identify submodular relationships; however, we go beyond submodularity to enable other geometric relationships and priors to be incorporated into the model. In [14], the focus is on characterizing when the problem becomes submodular.

The standard technique for solving non-submodular energies of this type is so-called Roof Duality (RD) [18] which is also proposed in [14]. However, the method is quite memory intensive and can be slow [7],[36]. We develop a Lagrangian dual approach to solve these non-submodular energies. The method is advantageous over roof duality; it uses almost half as much memory and we empirically show that it is faster. The speedup is likely and the memory usage is definitely due to the fact that, although both methods solve a min cut problem, our graph has only half the number of nodes compared to RD. At the same time our method computes a solution which is virtually identical to that of RD.

Another contribution is the evaluation of the optimization framework for medical segmentation problems. We apply our cardiac segmentation model on publicly available data sets and our optimization framework is compared to RD both in terms of memory and speed. This work is based on a previous workshop paper [40], which was focused on cardiac segmentation only. Here, we take a more general approach and demonstrate that the framework is applicable to other segmentation tasks as well. In addition, more qualitative and quantitative data measures are now given in the experiments.

II. MULTI-REGION FRAMEWORK

Before introducing the general framework we present an example of a construction using the framework as shown in Figure 2. Each model is built up as an energy function

where the minimizer of the energy function gives the desired segmentation.

Let $\mathcal{R}$ be the set of region labels excluding the background and let $\mathcal{P}$ be the set of voxel indices. Each voxel p should be assigned a region label $r \in \mathcal{R} \cup \{0\}$ where 0 is the background region. We introduce $\boldsymbol{x} \in \mathbf{B}^{|\mathcal{R}| \times |\mathcal{P}|}$, where $\mathbf{B} = \{0, 1\}$ and $\boldsymbol{x}$ is indexed as x_p^r with $r \in \mathcal{R}$ and $p \in \mathcal{P}$. Further, $\boldsymbol{x}^r$ represents all Boolean variables associated with region r and $\boldsymbol{x}_p$ represents all Boolean variables associated with voxel p. Each voxel in the image is represented by $|\mathcal{R}|$ Boolean variables, which will make it possible to directly encode geometric relationships between regions, like inclusion and exclusion.

Figure 2(c) shows the correspondence between r and $\boldsymbol{x}_p$ for the cardiac model. Here, the fact that for instance regions 2 and 3 should be contained in region 1 is encoded in the Boolean representation by the fact that the first Boolean variable is set to one. Similarly, region 4 is contained in both region 1 and region 2 and consequently, the first two Boolean variables are set to one.

The energy function to be minimized is:

$$E\left(\boldsymbol{x}\right) = D(\boldsymbol{x}) + V(\boldsymbol{x}) + W(\boldsymbol{x}), \tag{1}$$

whose three components are, in order, the unary terms, the pairwise terms (regularization) and the geometric interaction terms. For every voxel p, the unary terms introduce a cost for each labeling of $\boldsymbol{x}_p$:

$$D(\boldsymbol{x}) = \sum_{p \in \mathcal{P}} \sum_{r \in \mathcal{R}} D_p^r\left(x_p^r\right). \tag{2}$$

The pairwise terms use a connectivity $\mathcal{N}$ to favor smooth and correctly located boundaries:

$$V(\boldsymbol{x}) = \sum_{p,q \in \mathcal{N}} \sum_{r \in \mathcal{R}} V_{p,q}^r\left(x_p^r, x_q^r\right). \tag{3}$$

The geometric interaction terms associate a cost with labeling voxel p in region i with different labellings for voxel q in region j. These terms are used either to attract or repel different regions to each other:

$$W(\boldsymbol{x}) = \sum_{p,q \in \mathcal{N}} \sum_{\substack{i,j \in \mathcal{R} \\ i \neq j}} W_{p,q}^{i,j}\left(x_p^i, x_q^j\right). \tag{4}$$

For any voxel p, we denote the probability of this voxel to belong to region r by $P\left(x_p^r = 1\right)$.

A. Unary terms

The unary terms are constructed from the probability of each voxel belonging to any of the regions. We define

$$\mu_r(p) = -\log\left(\mathrm{P}\left(x_p^r = 1\right)\right), \tag{5}$$

for voxel p and region r.

Definition 1. We say that a region i is *parent* to region j if region j is forced to be contained inside i *directly*.

By directly we mean that if region j is forced to be contained inside region i via another region k, we only consider k as a parent to region j. Regions not forced to be contained inside

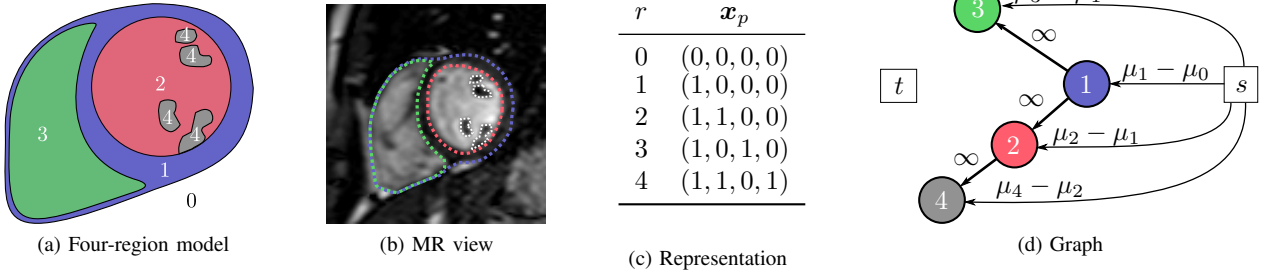(a) Four-region model (b) MR view (c) Representation (d) Graph

Figure 2. **(a)** A constructed short-axis view showing how the heart is modeled. Region 0 is the background, region 1 contains myocardium *and* the left and right ventricular cavities. Region 2 is the left ventricular cavity and region 3 the right ventricular cavity. Region 4 is the papillary muscles of the left ventricle. **(b)** An example of a slice from a short-axis image acquired with MRI where all four regions have been manually delineated. **(c)** The Boolean representation of the four regions reflect their geometric relationships as given in (a). **(d)** Graph construction for one voxel. The circled number corresponds to a vertex associated with the region number. The directed arrows are the directed edges in the graph. Best viewed in color.

any specific regions is defined to have the background $r = 0$ as parent.

As an example, consider our cardiac model in Figure 2. Region 4 has just one parent — region 2. Now consider any region r and let $\mathcal{G}_r$ denote the set of all parents to r, then we construct the unary term as

$$D_p^r \left(x_p^r \right) = x_p^r \left(\mu_r \left(p \right) - \sum_{g \in \mathcal{G}_r} \mu_g \left(p \right) \right), \qquad (6)$$

for all $p \in \mathcal{P}$ and $r \in \mathcal{R}$. Examples of these constructions are given in Figures 2 and 11. The reason this construction works is most easily explained through an example.

Example 2. Consider our cardiac model in Figure 2. According to (6) we end up with:

$$\sum_{r=1}^{4} D_p^r \left(x_p^r \right) = \quad x_p^4 \left(\mu_4 \left(p \right) - \mu_2 \left(p \right) \right) + x_p^3 \left(\mu_3 \left(p \right) - \mu_1 \left(p \right) \right)$$
$$+ \quad x_p^2 \left(\mu_2 \left(p \right) - \mu_1 \left(p \right) \right) + x_p^1 \left(\mu_1 \left(p \right) - \mu_0 \left(p \right) \right). \qquad (7)$$

Now consider a voxel assigned to region 4 from the model. We know that $x_p^1 = x_p^2 = x_p^4 = 1$ and $x_p^3 = 0$. It follows that

$$D_p^1 (1) + D_p^2 (1) + D_p^3 (0) + D_p^4 (1) =$$
$$1 \left(\mu_4 \left(p \right) - \mu_2 \left(p \right) \right) + 0 \left(\mu_3 \left(p \right) - \mu_1 \left(p \right) \right)$$
$$+ 1 \left(\mu_2 \left(p \right) - \mu_1 \left(p \right) \right) + 1 \left(\mu_1 \left(p \right) - \mu_0 \left(p \right) \right) =$$
$$\mu_4 \left(p \right) - \mu_0 \left(p \right). \qquad (8)$$

The reason this construction works is that Boolean variables with parents are linked to their parents by the geometric interaction term. The final cost for assigning a voxel to a region is added up like a telescopic sum resulting in $\mu_r - \mu_0$ for each region r.

B. Pairwise terms

The regularization weights are chosen differently for each region in a method related to the discussion in [17]. For each region i we choose the pairwise terms as:

$$V_{p,q}^r \left(x_p^r, x_q^r \right) = \frac{w_{p,q}}{1 + \beta \left(\mathrm{P} \left(x_p^r = r \right) - \mathrm{P} \left(x_q^r = r \right) \right)^2}, \qquad (9)$$
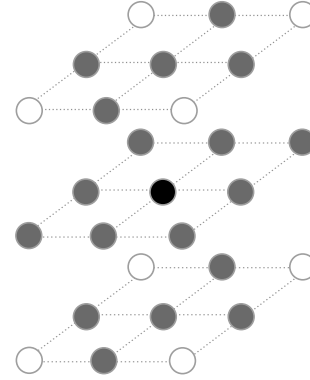

Figure 3. The 18-connected neighborhood visualized. The 18-connected neighborhood consists of all 18 nodes of distances at most $\sqrt{2}$ from the center node. In the image the center node is black and the 26 nodes closest to the center node in gray. A dashed line connect the nodes belonging to the same plane. The two types of edges have multipliers $w_{p,q}$ equal to 0.18714 and 0.16954, respectively, when not taking the anisotropic resolution into account.

where β can be used to tune the regularization. The neighborhood $\mathcal{N}$ for the regularization is in the experiments chosen as 18-connectivity, see Figure 3. The multipliers $w_{p,q}$ give different weights to different types of edges. One common choice is $w_{p,q} = 1/\operatorname{dist}(p, q)$; however, we instead use the arguably more correct way described in [9] based on solid angles. The fact that MRI has anisotropic resolution is very important to take into consideration both when calculating the distance between voxels and when using the method from [9].

C. Geometric interaction terms

We would both like some regions to be contained inside other regions, while forcing other regions apart. This is controlled by the geometric interaction terms. An energy function is *submodular* if it can be expressed as the sum of pairwise functions $E_{ij} \left(x_i, x_j \right)$ that satisfy:

$$E_{ij} \left(0, 0 \right) + E_{ij} \left(1, 1 \right) \leq E_{ij} \left(0, 1 \right) + E_{ij} \left(1, 0 \right).$$

The task to minimize such a function can be transformed into a min-cut problem on a s-t graph and is thus easily optimizable.

1) Submodular interaction terms: Suppose we would like region j to be *contained* inside region i. We then set

$$W_{p,p}^{i,j} \left(0, 1 \right) = \infty \text{ for all } p \in \mathcal{P}. \qquad (10)$$

This term is clearly submodular with our representation. It is also possible to enforce a margin between two regions by setting

$$W_{p,q}^{i,j}(0,1) = \infty \text{ for all } p \in \mathcal{P}, \tag{11}$$

where q is taken in some neighborhood $\mathcal{N}_p$ of p. As an example, let $\mathcal{N}_p$ be the 8-connected neighborhood of p. Now region j will not only be forced to be inside region i, it will be forced to be slightly smaller than region i. Using this we can enforce a margin such as a minimum distance between ventricles.

2) Non-submodular interaction terms: Similarly if we want region i to be *excluded* from region j we can set

$$W_{p,p}^{i,j}(1,1) = \infty \text{ for all } p \in \mathcal{P}. \tag{12}$$

This term is non-submodular. In some special cases the Boolean variables can be transformed in order to allow for a submodular construction with exclusion constraints [14]. However, this is possible for neither model in this paper. We will show in the next section that it is still possible to effectively optimize energy functions containing such non-submodular terms.

III. Solving the Optimization Problem

The standard approach for minimizing non-submodular functions of this type is to use roof duality. In this paper we present a new way to solve this using Lagrangian duality.

A. Solving using roof duality

The roof duality bound [5], [18] is a polynomial-time computable lower bound to the minimum of quadratic energy E. While a good lower bound can be interesting in many applications, for example, when using branch and bound, another property of roof duality is arguably even more useful: *persistency*. Each variable in the solution obtained via roof duality is equal to one of three possibilities: $\{0, 1, ?\}$. It is guaranteed that a global solution exists corresponding to the parts of the solution not equal to '?.' Therefore, the roof duality solution is said to be partially optimal. The unknown variables can then sometimes be assigned with a technique called *probing*, while still ensuring global optimality [7].

In [36] a complementary method to assign the unknown variables called *improve* was introduced. It works by randomly assigning variables to any given value and then run the algorithm again with these assumptions. By persistency results the old and the new solution can be combined in a way that is guaranteed to give a solution at least as good as the old one. Improve does not guarantee a globally optimal solution.

The fastest method of computing the roof dual for the energy function we are aware of is by using graph cuts [6]. The number of nodes in the graph will then be twice the number of variables, increasing the memory requirements significantly. In this paper, we use the implementation from [36].

One approach to solving our problem is to directly add the non-submodular terms to the energy function. It is then no longer graph-representable, but we can incorporate these energy terms using roof duality.

B. Using the Lagrangian dual

In this section we show how the problem of minimizing E can be reformulated using Lagrangian duality. The reformulated problem is then solved using the projected supergradient method.

a) Reformulating the problem: If we let $E'(\boldsymbol{x})$ be our energy without the non-submodular term, E' will be easy to minimize. Let $\boldsymbol{g}(\boldsymbol{x}) \leq 0$ be all non-submodular constraints. Adding these constraints gives us the new problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & E'(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}. \end{aligned} \tag{13}$$

This is the *primal* problem. We have now separated the easy part from the difficult non-submodular constraints. The primal problem can in principle be solved as an integer programming problem. However, this is not a tractable approach due to the large number of variables. Instead, we look at the *Lagrange dual* problem:

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \quad & d(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}} \left(\min_{\boldsymbol{x}} \left(E'(\boldsymbol{x}) + \boldsymbol{\lambda}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}) \right) \right) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq \boldsymbol{0}, \end{aligned} \tag{14}$$

where $d(\boldsymbol{\lambda})$ is the Lagrange dual function. Let $d^\star$ denote the optimal value for (14) and $p^\star$ the optimal value for (13). By weak duality we then have that $d^\star \leq p^\star$. For any solution to the dual problem we get a labeling $\boldsymbol{x}$ which we can project onto the set of feasible solutions. Using this labeling we get a primal energy p. If $d = p$ we know that we have found the global solution to problem and we can stop. If not we can look at the quantity $p - d$ which is known as the *duality gap*, which tells us how far away our solution $\boldsymbol{x}$ is from the *global* optimal solution. Since the energy function can be arbitrarily scaled by simply multiplying each cost with some constant we introduce the relative duality gap.

Definition 3. The *relative duality gap* is defined as $(p-d)/|p|$, were p and d are the currently best primal and dual energies for any iteration of the supergradient algorithm.

b) Solving the reformulated problem: The Lagrange dual function d is always concave. However, it is not differentiable in general, which precludes gradient ascent methods for its maximization. We can, however, use the projected supergradient method [33]. This method is similar to a gradient ascent method but has some key differences. Specifically, the method is easy to implement, but in general has worse convergence properties than first-order gradient-based methods. We refer the reader to [33], [38] for more details.

Definition 4. A *supergradient* to a function f at a point $\boldsymbol{x}_0$ is a vector $\boldsymbol{v}$ fulfilling $f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \leq (\boldsymbol{x} - \boldsymbol{x}_0)^T \boldsymbol{v}$, for every point $\boldsymbol{x}$.

Lemma 5 ([38]). *Let $\boldsymbol{\lambda}_0$ be given and let $\boldsymbol{x}^\star$ be the optimal solution to $d(\boldsymbol{\lambda}_0) = \min_{\boldsymbol{x}} \left(f(\boldsymbol{x}) + \boldsymbol{\lambda}_0^{\mathrm{T}} g(\boldsymbol{x}) \right)$. Then $g(\boldsymbol{x}^\star)$ is a supergradient to d at $\boldsymbol{\lambda}_0$.*

Proof: For any $\boldsymbol{\lambda}$ we have that

$$
\begin{aligned}
d\left(\boldsymbol{\lambda}\right) &\leq f\left(\boldsymbol{x}^{\star}\right) + \boldsymbol{\lambda}^{\mathrm{T}} g\left(\boldsymbol{x}^{\star}\right) \\
&= f\left(\boldsymbol{x}^{\star}\right) + \boldsymbol{\lambda}_0^{\mathrm{T}} g\left(\boldsymbol{x}^{\star}\right) + \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\right)^{\mathrm{T}} g\left(\boldsymbol{x}^{\star}\right) \\
&= \min_{\boldsymbol{x}}\left(f\left(\boldsymbol{x}\right) + \boldsymbol{\lambda}_0^{\mathrm{T}} g\left(\boldsymbol{x}\right)\right) + \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\right)^{\mathrm{T}} g\left(\boldsymbol{x}^{\star}\right) \\
&= d\left(\boldsymbol{\lambda}_0\right) + \left(\boldsymbol{\lambda} - \boldsymbol{\lambda}_0\right)^{\mathrm{T}} g\left(\boldsymbol{x}^{\star}\right).
\end{aligned}
$$

∎

To solve (14) we use the *projected supergradient method:*

1) Let $\boldsymbol{\lambda}_0$ be an initial guess of the optimal value of the concave function d.
2) In each step i a new possible solution $\boldsymbol{\lambda}_{i+1}$ is calculated as $\boldsymbol{\lambda}_{i+1} = \left[\boldsymbol{\lambda}_i + \tau_i \boldsymbol{v}_i\right]^+$, where $\boldsymbol{v}_i$ is any supergradient to d at $\boldsymbol{\lambda}_i$, τ_i is a step-length and $\left[\cdot\right]^+$ is a projection onto the feasible set $\{\boldsymbol{\lambda} \geq 0\}$.

For any inequality constraint $g\left(\boldsymbol{x}^{\star}\right) \leq 0$ we can use Lemma 5 and choose a supergradient for any given $\boldsymbol{\lambda}$ as $\boldsymbol{v} = g\left(\boldsymbol{x}^{\star}\right)$ where $\boldsymbol{x}^{\star}$ is the optimal solution of $d\left(\boldsymbol{\lambda}\right)$. We can now add any number of non-submodular constraint by simply introducing more Lagrange multipliers, one for each constraint.

In each step, the optimal solution $\boldsymbol{x}^{\star}$ for a chosen $\boldsymbol{\lambda}_i$ can be calculated via a minimum graph cut. Furthermore, as the edges will be very similar in each step, the graph structure can be reused, reducing the running time [24].

There are several ways of choosing the step-length. One family of step-lengths with favorable theoretical properties are the non-summable diminishing step-lengths.

Definition 6. A *non-summable diminishing* step-length τ_k satisfies

$$
\lim_{k \to \infty} \tau_k = 0, \quad \sum_{k=1}^{\infty} \tau_k = \infty. \tag{15}
$$

A simple example of a non-summable diminishing step-length is $\tau_i = 1/i$ where i is the iteration number. For non-summable and diminishing step-lengths and any given $\varepsilon > 0$ we are guaranteed that our solution $f(x)$ fulfills

$$
\lim_{k \to \infty} \left|\min_{i=1,\ldots,k} f\left(\boldsymbol{x}^{(i)}\right) - f\left(\boldsymbol{x}^{\star}\right)\right| < \varepsilon, \tag{16}
$$

where $\boldsymbol{x}^{\star} = \operatorname{argmin}_{\boldsymbol{x}} f\left(\boldsymbol{x}\right)$ [37].

However we have found the scheme employed by [38] to converge faster for our data. In this scheme each $\boldsymbol{x}_p$ is given its own step-length τ_p initialized at $\tau_p = 1$ and each time $g\left(\boldsymbol{x}_p\right)$ changes sign we set $\tau_p = \tau_p/2$. In this scheme each time a constraint in (13) changes from feasible to unfeasible the step-length for the variables $\boldsymbol{x}_p$ gets decreased. This is the scheme we have used for all experiments in this paper. Since supergradient methods do not guarantee improved value in each step the best solution thus far is always saved and once the *relative duality gap* is small enough the algorithm terminates.

Theorem 7 (Worst case scenario for supergradient method). *For any integer $k > 0$ and point $\boldsymbol{x}_0$ there exists a concave function $f\left(\boldsymbol{x}\right)$ with supergradients $v\left(\boldsymbol{x}\right)$ at point $\boldsymbol{x}$ such that any optimization scheme for f for which $\boldsymbol{x}_k$ is chosen as*
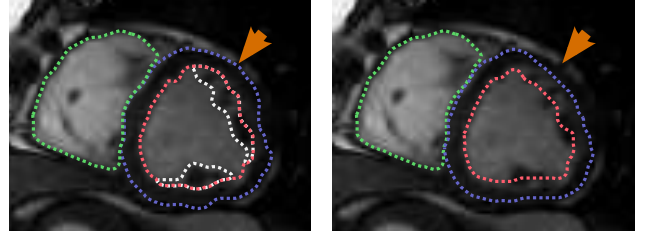

Figure 4. Example how modeling the papillary muscles improved both the segmentation of the left ventricle and the myocardium. (*left*) Complete model and (*right*) without modeling the papillary muscles.

$$
\boldsymbol{x}_k \in \boldsymbol{x}_0 + \operatorname{span}\left\{v\left(\boldsymbol{x}_o\right),\ldots,v\left(\boldsymbol{x}_{k-1}\right)\right\},
$$

needs $\mathcal{O}\left(1/\epsilon^2\right)$ *iterations to achieve*

$$
\left|f\left(\boldsymbol{x}_k\right) - f\left(\boldsymbol{x}^{\star}\right)\right| < \epsilon,
$$

where $\boldsymbol{x}^{\star} = \operatorname{argmin}_{\boldsymbol{x}} f\left(\boldsymbol{x}\right)$.

Proof: Theorems 3.2.1 and 3.2.2 in [33]. ∎

It should be noted that the rate of convergence in our applications is far better than the worst case of Theorem 7.

IV. Cardiac segmentation

The heart below the atrioventricular plane is modeled by four different regions as shown in Figures 2(a) and (b). The joint model describes both the geometry of the different regions and their appearances in the MR images. In the cardiac model, region 1 contains both region 2 and region 3. This is modeled by the use of geometric interaction terms as $W_{p,p}^{1,2}(0,1) = \infty$ and $W_{p,p}^{1,3}(0,1) = \infty$, for all $p \in \mathcal{P}$. Furthermore, the left ventricular papillary muscle must be inside the left ventricle. This is modeled as $W_{p,p}^{2,4}(0,1) = \infty$ for all $p \in \mathcal{P}$, see Figure 2(d). We also want to exclude region 2 from 3; that is, add terms of the form $W_{p,p}^{2,3}(1,1) = \infty$. These terms, however, become non-submodular and they can be handled by one of the methods in Section III. If we want to handle the non-submodular terms using Lagrangian duality we setup the primal optimization problem as:

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & E'(\boldsymbol{x}) \\
\text{subject to} \quad & \boldsymbol{x}^2 + \boldsymbol{x}^3 \leq 1,
\end{aligned}
$$

where $E'\left(\boldsymbol{x}\right)$ is our energy function without the non-submodular terms.

The unary terms construction as given in Figures 2(c) and (d) results in:

$$
\begin{aligned}
D_p^1(1) &= \mu_1(p) - \mu_0(p), \quad D_p^2(1) = \mu_2(p) - \mu_1(p), \quad (17) \\
D_p^3(1) &= \mu_3(p) - \mu_1(p), \quad D_p^4(1) = \mu_4(p) - \mu_2(p),
\end{aligned}
$$

and $D_p^r(0) = 0$ for all $r \in \mathcal{R}$ and $p \in \mathcal{P}$.

For the heart model we split the spatial probability into four categories: left ventricle, right ventricle, myocardium and background. Similarly the intensity is split into three categories: blood, muscle and background. The probability for each region is then calculated with the assumption that the spatial and intensity distributions are independent. An example of the
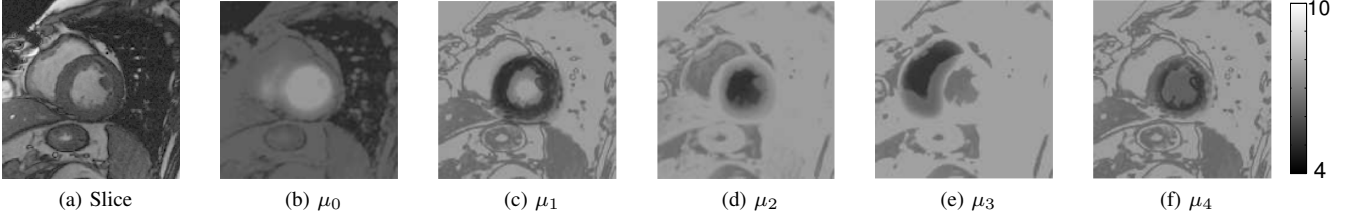
| (a) Slice | (b) μ_0 | (c) μ_1 | (d) μ_2 | (e) μ_3 | (f) μ_4 |

Figure 5. Example of μ_r for the slice shown in (a). Recall that $\mu_r(p) = -\log\left(\mathrm{P}\left(x_p^r = 1\right)\right)$. A lower intensity corresponds to higher probability.
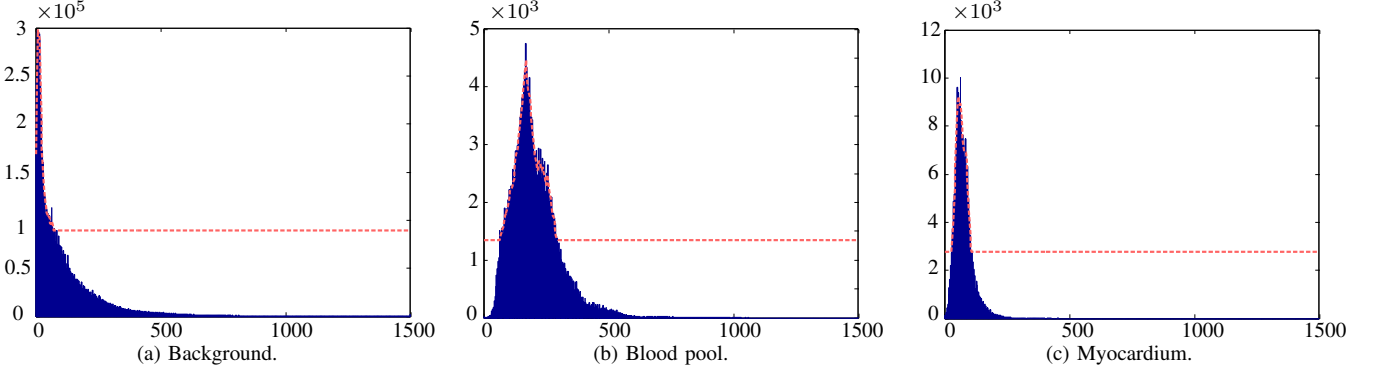

| (a) Background. | (b) Blood pool. | (c) Myocardium. |

Figure 6. Estimated intensity distributions for the three classes of intensity. Every example in the training data is added together and put into histogram of 500 bins from 0 to 1500 shown as the blue bars. The resulting 1D data is smoothed using a Gaussian kernel. For the smoothed data the bins are sorted by occurrence and the bin where the cumulative sum is 90% of the total is chosen as threshold level giving the dashed red line. The final distributions are normalized versions of the three lines.

final μ_r's can be found in Figure 5. The spatial distribution is estimated by first resizing each image in the training data to the same size by bilinear interpolation. Then a binary mask is constructed for each category. The masks are enlarged and smoothed and then they are all added together constructing the final probability mask. The intensity distribution for each region is estimated by collecting all intensities from the examples in the training data. The histogram of intensities is then smoothed and a distribution is constructed, see Figure 6. For both the location and intensity probability a lowest probability is set, in order to capture occurrences unseen in the training data. The user selects which slices to be segmented and selects a center point of the right and left ventricle in *one* slice. The two center points are used to roughly align the hearts in order to get good spatial statistics. The algorithm can handle slices lacking any of the regions. Badly captured MRIs are identified by looking at the distribution of the intensities. If there are multiple peaks in the histogram close to each other for the lower intensities, the image is assumed to be too bright and the intensity distribution is shifted to fit an average histogram.

In all ground truth data we have come across, only the left ventricular epicardium is delineated. In our model we do not have this restriction — we segment the full myocardium. In order to compare our results with the ground truth we must remove all myocardium which is not part of the left ventricular epicardium. To do this, the thickness of the septum is approximated as the shortest distance between the left and right ventricles in the resulting segmentation. Then the outlying myocardium is removed based on this thickness approximation, cf. Figure 2b(a). We also assume that the left ventricle and the myocardium are convex. The resulting segmentation is taken as the convex hull in each slice.

The regularization can sometimes make the segmentation miss the apical slice. By user input we know which slices the left and right ventricles are contained in and it would be wasteful to throw this information away. We utilize the user input by naturally extending the segmentation into the apical slice. This is done by taking the segmentation from another slice, shrinking it slightly and inserting at the bottom.

A. Experiments

The segmentation is only performed on the slices of the heart which are fully below the atrioventricular plane. The quality of the segmentation is measured by the dice metric.

Definition 8. The dice metric is given by $2\left|A \cap B\right| / \left(|A| + |B|\right)$, where A and B are the ground truth and the computed segmentations, respectively.

The algorithm is evaluated on two data sets: Lund and Sunnybrook. Each data set is trained and evaluated separately.

Lund consists of cine short-axis steady state free precession MR images of 62 healthy normals captured on a Philips Interera CV 1.5T with a five channel cardiac synergy coil. Each heart has the left and right ventricular endocardium and the left ventricular epicardium manually delineated by an expert. The data set is split into two equally sized parts, one used for training and one used for evaluation. Results are given in Table I(a) and an example segmentation in Figure 7. We also evaluate three clinical parameters: the left ventricular mass has an error of $15.6_{\pm 11.5}$ g, the left and right ventricular ejection fraction errors are $5.6_{\pm 2.9}\%$ and $7.1_{\pm 5.2}\%$, respectively.
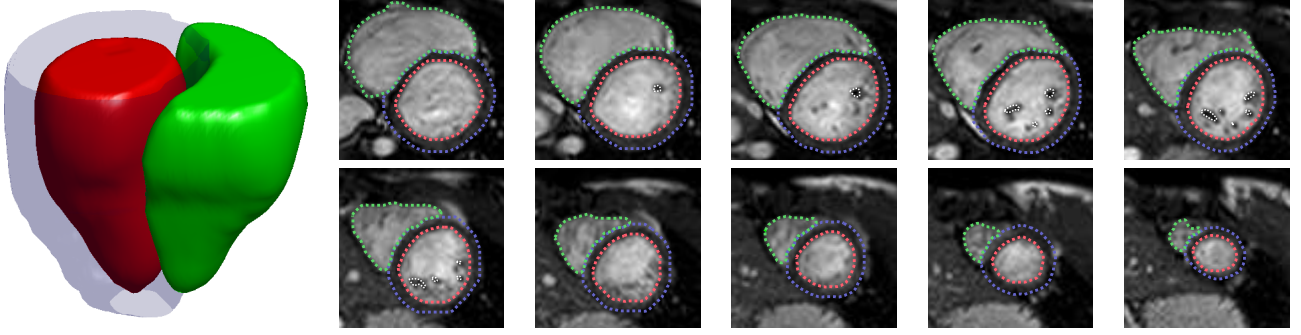
Figure 7. Example segmentation from Lund, (left) 3D rendering and (right) 10 slices

	End systole			End diastole		
	LV endo.	LV epi.	RV	LV endo.	LV epi.	RV
Multi-region	$0.87{\pm}0.05$	$0.88{\pm}0.05$	$0.80{\pm}0.11$	$0.96{\pm}0.02$	$0.93{\pm}0.03$	$0.91{\pm}0.07$
Separately	$0.47{\pm}0.25$	$0.86{\pm}0.04$	$0.42{\pm}0.14$	$0.62{\pm}0.12$	$0.90{\pm}0.03$	$0.57{\pm}0.14$

Table I
RESULTS IN THE DICE METRIC FOR LUND REPORTED AS MEAN $\pm$ ONE STANDARD DEVIATION. NOTE THAT THE MULTI-REGION MODEL HAS A HUGE INFLUENCE ON THE SEGMENTATION RESULTS.

We also compare our method to a simplified version where we run the segmentation for each region separately, see Table I(b). Without the complete multi-region model, the localization of the ventricles becomes very difficult and the blood pools are often overestimated. A few typical examples where the multi-region model improves the segmentation are given in Figures 4 and 8.

Sunnybrook consists of 30 patients with different heart diseases and is split up into two equally sized parts, one for training and one for evaluation. The data set was used in the 2009 MICCAI segmentation challenge [1]. Sunnybrook lacks ground truth for the right ventricles, so this was manually constructed by a non-expert. Therefore, this ground truth was only used for training and not for evaluation. The results given by the evaluation code used in the challenge are given in Table III along with results from competing methods. The evaluation in the challenge calculates the dice metric per slice and averages over all slices.

The small training data of Sunnybrook gives our method a disadvantage as there are just 15 hearts spanning over three different diseases and one group of normals. Image-driven methods do not suffer from the small training set as they do not need to be trained. The limited number of training examples impedes the model since there are too few examples of variation in shape for each disease and the normals. The intensity model is less effected by this but would still benefit from a larger training set. Note that all diseases and normals are covered by one model.

Our model was also optimized with roof duality (RD). If RD was unable to label all variables, then the methods "probe" and "improve" were used to obtain a complete labelling. We call the method *RD-I* if first RD is calculated and then improve is used to label the last variables and *RD-P* if probing is used to label the unlabelled variables. For RD-I and our method we used the same termination criterion: either the relative duality

gap was smaller than 10^{-4} or a maximum of 25 iterations. RD-P terminated either if all variables have been labelled or after a maximum of 12 hours running time. If some variables still was unlabelled after 12 hours they were set to 0. Our method was faster and used less memory than RD-I and RD-P. The final results for all the optimization methods in terms of quality of segmentation are virtually identical, see Table II. In particular, all methods achieved small duality gaps. Only RD-P encountered some problems on the Sunnybrook dataset, where the larger duality gap is a result of 3 hearts that were not completely labelled after 12 hours of probing. The progress of the duality gap and dice over time for the different optimization methods is depicted in Figure 9. In Figure 10, the progress of probing over time is shown. It takes a relatively large computational effort for probing to make a difference.

Our method found a globally optimal solution for 52% of the hearts, and for the other hearts we can from the very small relative duality gap be certain that the method found a solution close to the global optimum. For 4 out of a total of 46 hearts the probing took more than 12 hours and we terminated the calculations after that time. This highlights the problem with probing - there is no real guarantee that the computations will be done within a reasonable time; on some problem instances we had probing running for several weeks without returning a complete solution.

It is possible to extend the cardiac model to also include papillary muscles in the right ventricle; we need only to introduce one more variable per voxel. If we let $\mu_5(p) = \log P(x_p = 5)$ for this new region and follow the notation in Figure 2(d) we need only to add one vertex corresponding to the new variable and two edges: one s-t edge with value $\mu_5 - \mu_3$ and one edge going from region 5 to region 3 with ∞ weight. Initial experiments gave worse results for both the right ventricle and myocardium segmentation with the added region. The new region had a tendency to overflow into the septum since
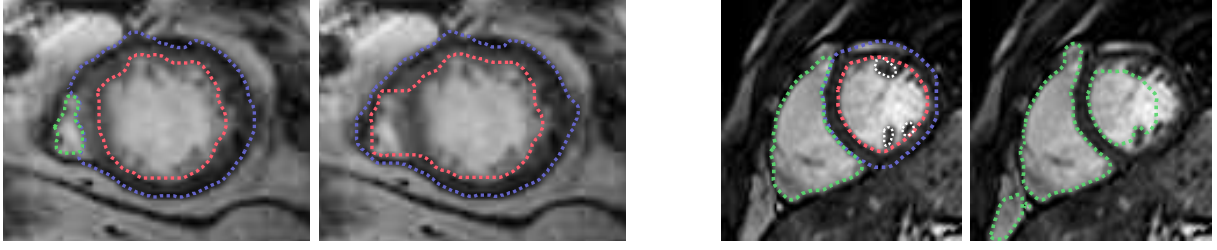
(a) (*left*) Complete model and (*right*) without modeling the right ventricle.

(b) (*left*) Complete model and (*right*) only the right ventricle.

Figure 8. Examples of how modeling multiple regions improve the segmentation of the ventricular epi- and endocardium. The color scheme is the same as in Figure 2b.
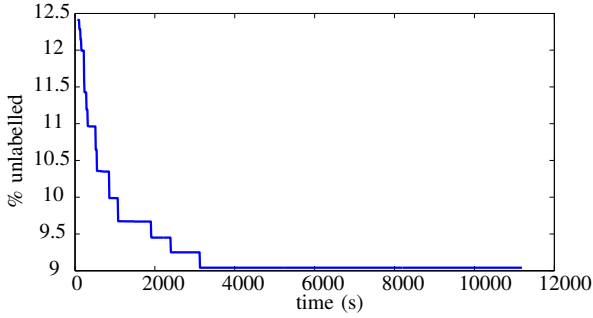

Figure 10. The progress of probing as a function of time averaged over all hearts. Most improvements occur within the first hour.

this would give region 3 a rounder shape giving a lower regularization cost, and therefore it has not been incorporated in the model.

The Lund data set was manually delineated using both short- and long-axis images. For a number of hearts the most basal slice for the short-axis images containing the left ventricular cavity also cut through to the atrium. For these slices it was hard or even impossible to even manually delineate the left ventricle solely based on information from the short-axis images. When the ground truth was produced, long-axis images were used to properly segment them. It would be desirable for our algorithm to incorporate information from long-axis images as well so we also could handle these few slices.

V. Lung segmentation

Our second application is the segmentation of lungs in a full-body X-ray CT scan. The model is shown in Figure 11 and uses four regions: the body (region 1), the two lungs (regions 2,3) and the heart together with the throat (region 4). Regions 2, 3 and 4 are all forced to be contained inside region 1 by adding the terms $W_p^{1,2}(0,1) = \infty$, $W_p^{1,3}(0,1) = \infty$ and $W_p^{1,4}(0,1) = \infty$ for all $p \in \mathcal{P}$. The major difference to the cardiac model is that we now need to enforce more than one separation of regions, that is,

$$\min_{\boldsymbol{x}} \quad E'(\boldsymbol{x})$$
$$\text{subject to} \quad \boldsymbol{x}^2 + \boldsymbol{x}^3 + \boldsymbol{x}^4 \leq 1. \tag{18}$$

Alternatively, the three-variable constraint could equivalently be replaced with three constraints of the same type as in the

previous section:

$$\boldsymbol{x}^2 + \boldsymbol{x}^3 \leq 1 \tag{19}$$
$$\boldsymbol{x}^2 + \boldsymbol{x}^4 \leq 1$$
$$\boldsymbol{x}^3 + \boldsymbol{x}^4 \leq 1.$$

These two sets of constraints perform almost identically in our experiments, but, obviously, the second set requires three times as many dual variables.

The unary terms construction can be seen in Figure 11(d). We get:

$$D_p^1(1) = \mu_1(p) - \mu_0(p), \quad D_p^2(1) = \mu_2(p) - \mu_1(p), \tag{20}$$
$$D_p^3(1) = \mu_3(p) - \mu_1(p), \quad D_p^4(1) = \mu_4(p) - \mu_1(p),$$

and $D_p^r(0) = 0$ for all $r \in \mathcal{R}$ and $p \in \mathcal{P}$.

The user gives ground truth seeds only in one slice of the data as shown in Figure 11(b). The background is removed by thresholding on an intensity level between the seeds given from the background and the body. The seeds are then used to build intensity histograms for the five regions which are used to estimate the intensity distribution. We do not build any kind of spatial statistics; we simply give the probability for the left to be to the right of the side each slice as a fading gradient with respect to the center and vice versa for the left lung. We construct the unary terms using the approximated intensity distributions, see Figure 13 for examples.

A. Experiments

We test our algorithm on a full-body X-ray CT data set with seed as shown in Figure 11. A sample result from a few slices can be seen in Figure 12. The running time for roof duality is 39 seconds and for our method 29 seconds. Both methods give exactly same solution.

In the current implementation we used two different general-purpose max-flow implementations [10], [15]. The performance of the two algorithms was quite similar. One thing not taken into account is the fact that the structure of the graph is highly repetitive. For instance, all geometric interaction terms are equal and they need not be explicitly stored in the graph. A specialized-purpose solver for this problem could lead to a large reduction of memory requirements.

| Method	Memory (MB)		Running time (s)		Relative duality gap		Dice (average)	
	Sunnybrook	Lund	Sunnybrook	Lund	Sunnybrook	Lund	Sunnybrook	Lund
Our	2727 ± 680	2103 ± 788	46 ± 27	30 ± 27	0.00054 ± 0.0013	0.00054 ± 0.0021	0.888 ± 0.0484	0.892 ± 0.0815
RD-I	5038 ± 985	3913 ± 1407	135 ± 165	80 ± 113	0.00016 ± 0.00034	0.00049 ± 0.0021	0.888 ± 0.0485	0.892 ± 0.0816
RD-P	5041 ± 1014	3949 ± 1402	6109 ± 12451	1934 ± 7984	0.0011 ± 0.0030	0.00056 ± 0.0021	0.888 ± 0.0484	0.892 ± 0.0825

Table II

MEMORY CONSUMPTION OF THE OPTIMIZATION IN MEGABYTE AND THE RELATIVE DUALITY GAP. THE RESOLUTION OF THE DATA FED TO THE METHOD WAS FOR SUNNYBROOK ON AVERAGE $146 \times 146 \times 10 \times 2$ VOXELS AND FOR LUND ON AVERAGE $126 \times 126 \times 10 \times 2$ VOXELS, THE THIRD DIMENSION, THE NUMBER OF SLICES, VARIED FROM HEART TO HEART. COMPARING EACH PROBLEM INSTANCE, INSTEAD OF THE TOTAL MEAN, BOTH VERSIONS OF RD USES ≈ 1.9 TIMES MORE MEMORY THAN OUR METHOD. DICE WAS TAKEN AS AN AVERAGE OVER PARTS WHERE THE DATASET HAD GROUND TRUTH.

Method	Dice		LV Mass (g)	LV ejection fraction $(\%)$
	LV endo.	LV epi.		
Our	0.86 ± 0.05	0.92 ± 0.02	27.1 ± 28.3	12.5 ± 8.7
Mark et al. [30]	0.86 ± 0.04	0.93 ± 0.01	$23\pm?$	$14\pm?$
Lu et al. [29]	0.89 ± 0.03	0.94 ± 0.02	21.6 ± 14.6	8.08 ± 5.06
Wijnhout et al. [41]	0.89 ± 0.03	0.93 ± 0.01	28.7 ± 18.7	7.02 ± 4.78
Casta et al. [12]	?	$0.93\pm?$	†	?
O'Brien et al. [34]	$0.81\pm?$	$0.91\pm?$	?	?
Constantinides et al. [13]	0.89 ± 0.04	0.92 ± 0.02	†	†
Huang et al. [21]	0.89 ± 0.04	0.94 ± 0.01	?	?
Jolly [23]	0.88 ± 0.04	0.93 ± 0.02	31.8 ± 17.7	8.35 ± 5.78

Table III

RESULTS FOR SUNNYBROOK. "?" MEANS NOT REPORTED IN THE CORRESPONDING PAPER. "†" MEANS THAT THE RESULT IS NOT DIRECTLY COMPARABLE. MASS AND EJECTION FRACTION IS REPORTED AS DIFFERENCE BETWEEN MANUAL AND AUTOMATIC VALUE.
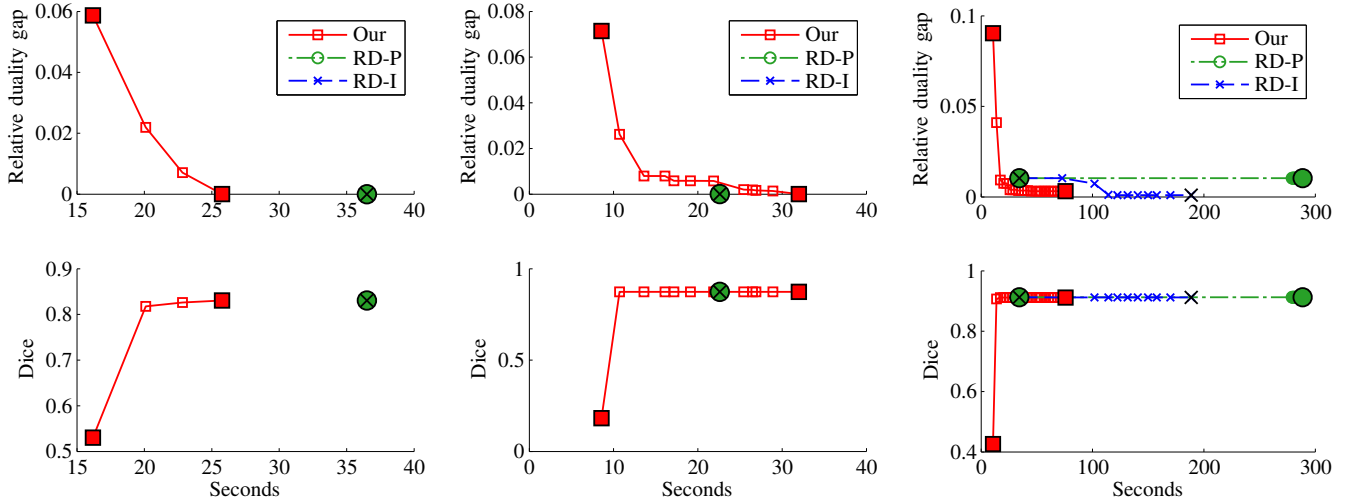
(a) Our method converges fast to a solution with small duality gap (70%).

(b) RD-P and RD-I is effective (6%).

(c) RD-P is slow while our method and RD-I converge relatively fast. (24%).

Figure 9. The typical progress for all methods, each iteration is shown by a marker with the first and last iteration shown as a larger marker. For all methods the resulting dice are virtually identical. The results can be divided into three different categories. In a) all iterations of our method are done before the initial RD calculations are completed. This happened for 70% of the hearts. In b) RD managed to label all nodes before our method terminated, which occurred for 6% of the hearts. In c) RD was unable to label all nodes and RD-P converged very slowly, which happened for 24% of the hearts. The top images show relative duality gap and the bottom images show dice as a function of time for the three methods. The first and last iterations for each method are highlighted and for RD-P we limited the number of iterations to 25.
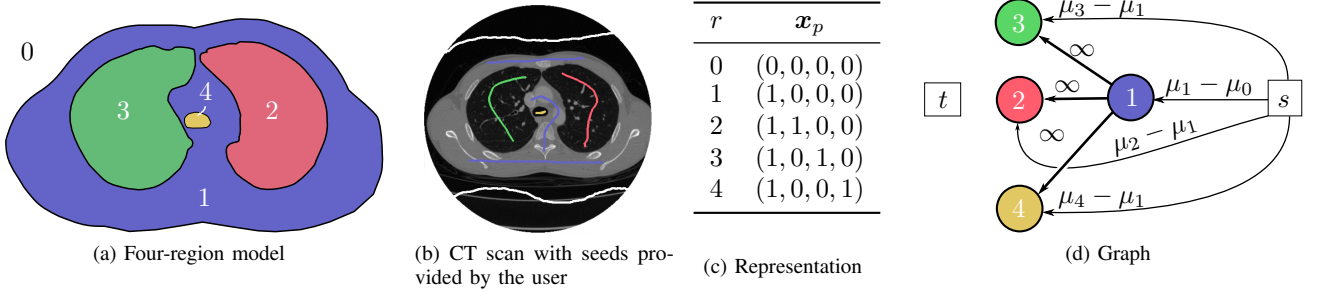
(a) Four-region model

(b) CT scan with seeds provided by the user

(c) Representation

r	$\boldsymbol{x}_p$
0	$(0,0,0,0)$
1	$(1,0,0,0)$
2	$(1,1,0,0)$
3	$(1,0,1,0)$
4	$(1,0,0,1)$

(d) Graph

Figure 11. **(a)** A diagram showing the model used for lung segmentation. Region 0 is the background, region 1 the body, regions 2 and 3 are the right and left lungs, respectively, and region 4 is the throat. **(b)** The seeds in one slice used for the segmentation. In a clinical setting, these are provided by a physician. **(c)** The Boolean representation of the four regions. **(d)** Graph construction for one voxel, showing the geometrical relationships. Best viewed in color.
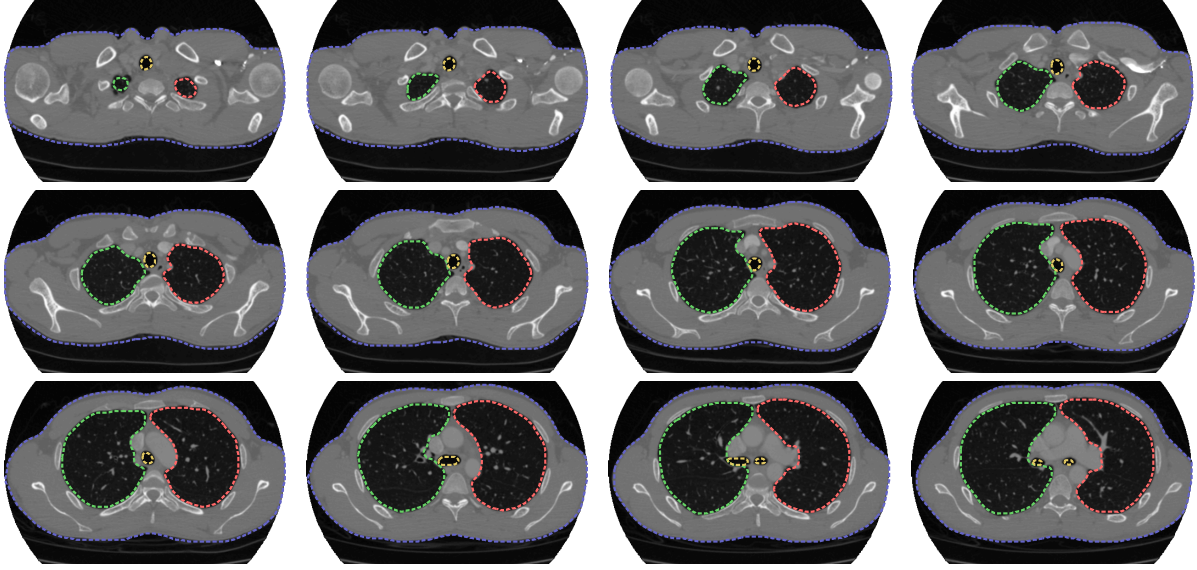
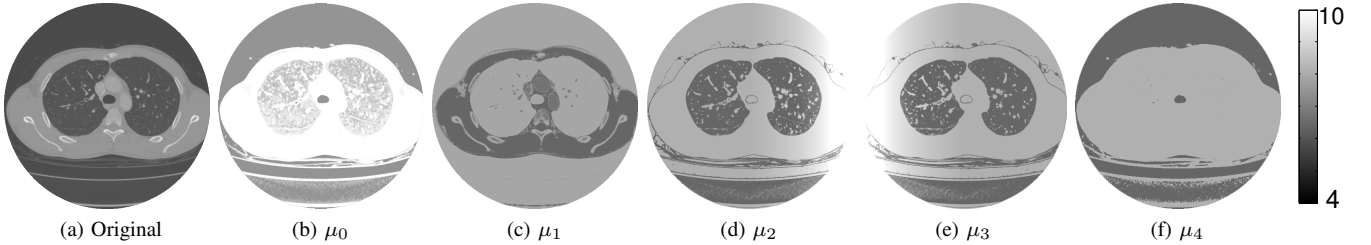Figure 12. Sample results from the segmentation, the same color coding as in Figure 11 is used.

(a) Original (b) μ_0 (c) μ_1 (d) μ_2 (e) μ_3 (f) μ_4

Figure 13. An example slice from the data set with the 5 calculated unary terms for this slice.

VI. CONCLUSIONS

Based on the experimental results we can draw the following three conclusions:

1) We have demonstrated that a multi-region model achieves significantly better results, all else being equal, than segmenting the regions one at a time (Figures 1,4 and 8). Enforcing geometric constraints, and more generally, incorporating prior information into the model, result in qualitative improvements. This is not always captured by quantitative measures such as the dice metric.

2) The optimization method based on Lagrangian duality outperforms roof duality, both in terms of speed and memory consumption.

3) Application of the multi-region framework for cardiac segmentation achieves results on par with dedicated LV methods on a publicly available data base. There are fine-tunings one can make to improve performance, better unary term and better parameter choices. However we still find these results encouraging.

The model can simply be modified to fit other medical imaging tasks by adding and removing regions as demonstrated by the small difference between the heart and the lung model.

REFERENCES

[1] *MICCAI Cardiac MR Left Ventricle Segmentation Challenge*, 2009, http://smial.sri.utoronto.ca/LV_Challenge/Home.html. IV-A

[2] B. Appleton and H. Talbot, "Globally minimal surfaces by continuous maximal flows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 106–118, 2006. I-A

[3] S. Armato III and W. Sensakovic, "Automated lung segmentation for thoracic ct: Impact on computer-aided diagnosis," *Academic Radiology*, vol. 11, no. 9, pp. 1011–1021, 2004. I-A

[4] A. Blake, P. Kohli, and C. Rother, Eds., *Markov Random Fields for Vision and Image Processing*. MIT Press, 2011. I-A

[5] E. Boros and P. Hammer, "Pseudo-boolean optimization," *Discrete Appl. Math.*, vol. 123, pp. 155–225, 2002. III-A

[6] E. Boros, P. Hammer, and X. Sun, "Network flows and minimization of quadratic pseudo-boolean functions," RUTCOR RRR 17-1991, Tech. Rep., 1991. III-A

[7] E. Boros, P. Hammer, and G. Tavares, "Preprocessing of unconstrained quadratic binary optimization," RUTCOR RRR 10-2006, Tech. Rep., 2006. I-B, III-A

[8] Y. Boykov and M. Jolly, "Interactive organ segmentation using graph cuts," in *Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Pittsburgh, USA: Springer, 2000, pp. 147–175. I-A

[9] Y. Boykov and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts," in *Int. Conf. Computer Vision*, Nice, France, 2003, pp. 26–33. II-B

[10] ——, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004. V-A

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001. I-A

[12] C. Casta, P. Clarysse, J. Schaerer, and J. Pousin. (2009) Evaluation of the Dynamic Deformable Elastic Template model for the segmentation of the heart in MRI sequences. [Online]. Available: http://hdl.handle.net/10380/3072 IV-A

[13] C. Constantinides, Y. Chenoune, N. Kachenoura, E. Roullot, E. Mousseaux, A. Herment, and F. Frouin. (2009) Semi-automated cardiac segmentation on cine magnetic resonance images using GVF-Snake deformable models. [Online]. Available: http://hdl.handle.net/10380/3108 IV-A

[14] A. Delong and Y. Boykov, "Globally optimal segmentation of multi-region objects," in *Int. Conf. Computer Vision*, Kyoto, Japan, 2009. I-B, II-C2

[15] A. Goldberg, S. Hed, H. Kaplan, R. Tarjan, and R. Werneck, "Maximum flows by incremental breadth-first search," in *European Symposium on Algorithms, ALGO ESA*, Saarbrücken, Germany, 2011, pp. 457–468. V-A

[16] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006. I-A

[17] L. Grady and M. Jolly, "Weights and topology: A study of the effects of graph construction on 3d image segmentation," in *Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, New York City, USA, 2008, pp. 153–161. II-B

[18] P. Hammer, P. Hansen, and B. Simeone, "Roof duality, complementation and persistency in quadratic 0-1 optimization," *Math. Programming*, vol. 28, no. 2, pp. 121–155, 1984. I-B, III-A

[19] E. Heiberg *et al.*, "Design and validation of Segment- freely available software for cardiovascular image analysis," *BMC Medical Imaging*, vol. 10, no. 1, p. 1, 2010. VI

[20] S. Hu, E. Hoffman, and J. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric x-ray ct images," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 6, pp. 490 –498, june 2001. I-A

[21] S. Huang, J. Liu, L. Lee, S. Venkatesh, L. Teo, C. Au, and W. Nowinski. (2009, 08) Segmentation of the Left Ventricle from Cine MR Images Using a Comprehensive Approach. [Online]. Available: http://hdl.handle.net/10380/3121 IV-A

[22] H. Ishikawa, "Exact optimization for markov random fields with convex priors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1132–1139, 2003. I-A

[23] M. Jolly. (2009) Fully Automatic Left Ventricle Segmentation in Cardiac Cine MR Images Using Registration and Minimum Surfaces. [Online]. Available: http://hdl.handle.net/10380/3114 IV-A

[24] P. Kohli and P. Torr, "Dynamic graph cuts for efficient inference in markov random fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007. III-B0b

[25] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1568–1583, 2006. I-A

[26] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1436–1453, 2007. I-A

[27] X. Lin, B. Cowan, and A. Young, "Model-based graph cut method for segmentation of the left ventricle," in *Engineering in Medicine and Biology Society (IEEE-EMBS)*, Shanghai, China, 2005. I-A

[28] M. Lorenzo-Valdés *et al.*, "Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm," *Medical Image Analysis*, vol. 8, no. 3, pp. 255–265, 2004. I-A

[29] Y. Lu, P. Radau, K. Connelly, A. Dick, and G. Wright. (2009) Automatic Image-Driven Segmentation of Left Ventricle in Cardiac Cine MRI. [Online]. Available: http://hdl.handle.net/10380/3109 IV-A

[30] L. Marak, J. Cousty, L. Najman, and H. Talbot. (2009) 4D Morphological segmentation and the MICCAI LV-segmentation grand challenge. [Online]. Available: http://hdl.handle.net/10380/3085 IV-A

[31] C. McIntosh and G. Hamarneh, "Medial-based deformable models in non-convex shape-spaces for medical image segmentation using genetic algorithms," *IEEE transactions on medical imaging*, 2011. I-A

[32] S. Mitchell, B. F. Lelieveldt, R. van der Geest, H. G. Bosch, J. Reiber, and M. Sonka, "Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images," *IEEE Trans. Medical Imaging*, vol. 20, no. 5, pp. 415–423, 2002. I-A

[33] Y. Nesterov, *Introductory Lectures on Convex Optimization*. Kluwer Academic Publishers, 2004. III-B0b

[34] S. O'Brien, O. Ghita, and P. Whelan. (2009) Segmenting the Left Ventricle in 3D Using a Coupled ASM and a Learned Non-Rigid Spatial Model. [Online]. Available: http://hdl.handle.net/10380/3110 IV-A

[35] N. Paragios, "A variational approach for the segmentation of the left ventricle in cardiac image analysis," *Int. Journal Computer Vision*, vol. 50, no. 3, pp. 345–362, 2002. I-A

[36] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer., "Optimizing binary MRFs via extended roof duality." in *Conf. Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, 2007, pp. 1–8. I-B, III-A

[37] L. X. Stephen Boyd and A. Mutapcic. (2003) Subgradient methods. [Online]. Available: http://www.stanford.edu/class/ee392o/subgrad_method.pdf III-B0b

[38] P. Strandmark and F. Kahl, "Parallel and Distributed Graph Cuts by Dual Decomposition," in *Conf. Computer Vision and Pattern Recognition*, San Francisco, USA, 2010, pp. 2085–2092. III-B0b, 5, III-B0b

[39] S. Ukil and J. Reinhardt, "Anatomy-guided lung lobe segmentation in x-ray ct images," *IEEE Trans. Medical Imaging*, vol. 28, no. 2, pp. 202–214, 2009. I-A

[40] J. Ulén, P. Strandmark, and F. Kahl, "Optimization for multi-region segmentation of cardiac MRI," in *MICCAI Workshop on Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges*, Toronto, Canada, 2011, pp. 129–138. I-B

[41] J. Wijnhout, D. Hendriksen, H. van Assen, and R. van der Geest. (2009) LV challenge LKEB contribution: Fully automated myocardial contour detection. [Online]. Available: http://hdl.handle.net/10380/3115 IV-A