ACTIVE MEAN FIELDS FOR PROBABILISTIC IMAGE SEGMENTATION: CONNECTIONS WITH CHAN-VESE AND RUDIN-OSHER-FATEMI MODELS

MARC NIETHAMMER*, KILIAN M. POHL†, FIRDAUS JANOOS‡, AND WILLIAM M. WELLS III§

Abstract. Segmentation is a fundamental task for extracting semantically meaningful regions from an image. The goal of segmentation algorithms is to accurately assign object labels to each image location. However, image-noise, shortcomings of algorithms, and image ambiguities cause uncertainty in label assignment. Estimating the uncertainty in label assignment is important in multiple application domains, such as segmenting tumors from medical images for radiation treatment planning. One way to estimate these uncertainties is through the computation of posteriors of Bayesian models, which is computationally prohibitive for many practical applications. On the other hand, most computationally efficient methods fail to estimate label uncertainty. We therefore propose in this paper the Active Mean Fields (AMF) approach, a technique based on Bayesian modeling that uses a mean-field approximation to efficiently compute a segmentation and its corresponding uncertainty. Based on a variational formulation, the resulting convex model combines any label-likelihood measure with a prior on the length of the segmentation boundary. A specific implementation of that model is the Chan–Vese segmentation model (CV), in which the binary segmentation task is defined by a Gaussian likelihood and a prior regularizing the length of the segmentation boundary. Furthermore, the Euler–Lagrange equations derived from the AMF model are equivalent to those of the popular Rudin-Osher-Fatemi (ROF) model for image denoising. Solutions to the AMF model can thus be implemented by directly utilizing highly-efficient ROF solvers on log-likelihood ratio fields. We qualitatively assess the approach on synthetic data as well as on real natural and medical images. For a quantitative evaluation, we apply our approach to the `icgbench` dataset.

Key words. Segmentation, mean-field approximation, Rudin-Osher-Fatemi model, Chan-Vese model

AMS subject classifications.

1. Introduction.

Image segmentation approaches rarely provide measures of segmentation label uncertainty. In fact, most existing and probabilistically-motivated segmentation approaches only compute the *maximum a posteriori* (MAP) solution [34, 35, 8, 20, 44]. Using these models to segment ambiguous boundaries is troublesome especially for applications where confidence in object boundaries impacts analysis. For example, many radiation treatment plans base dose distribution on the boundaries of tumors segmented from medical images with low contrast [37]. This can be problematic, as segmentation variability can have a substantial effect on radiation treatment; Martin et al. [37] report that such variability caused mean observer tumor control probability (*i.e.*, the probability to control or eradicate a tumor at a given dose) to range from (22.6 $\pm$11.9)% to (33.7 $\pm$ 0.6)% between six participating physicians in a study of intensity-modulated radiation therapy (IMRT) of 4D-CT-based non-small cell lung cancer radiotherapy. The precision of the planning could be improved around highly-confident tumor boundaries [37, 30] thereby reducing the risk of damaging healthy tissue in those areas. As significant information about label uncertainty is contained in the posterior *distribution*, it is natural to go beyond determining

*University of North Carolina at Chapel Hill, Department of Computer Science and Biomedical Research Imaging Center (BRIC), (mn@cs.unc.edu).

†Center for Health Sciences, SRI International, (kilian.pohl@sri.com).

‡Two Sigma Investments, (firdaus.janoos@gmail.com).

§Harvard Medical School and Brigham and Women's Hospital, (sw@bwh.harvard.edu).

a MAP solution and instead to either compute the posterior distribution itself or a computationally efficient approximation.

This paper develops such a method for an efficient approximation of the posterior distribution on labels. Furthermore, it connects this method to the Rudin-Osher-Fatemi (ROF) model for image-denoising [51, 57, 3] and previously existing level-set segmentation approaches [42], in particular the Chan-Vese segmentation model [15]. Due to these connections we can (i) make use of the efficient solvers for the ROF model to approximate the posterior distribution on labels and (ii) compute the solution to the Chan-Vese model through the MAP realization of our approximation to the posterior distribution, *i.e.*, our model is more general and subsumes the Chan-Vese model. In contrast to the implicit style of active-contour methods that represent labels by way of zero level-sets, such as the classical formulation of the Chan-Vese model, we use a dense logit ("log odds"), representation of label probabilities. This is akin to the convex approaches for active contours [2], but in a probabilistic formulation.

1.1. Motivations. Beyond optimal labelings, posterior distributions on labelings offer some advantages. For example, in many instances, one wishes to obtain information about segmentation confidence; or in change detection, distributions can help to determine whether an observed apparent change may be due to chance. Furthermore, probabilistic models on latent label fields can be useful for constructing more ambitious systems that, e.g., perform simultaneous segmentation and atlas registration [49]. However, the computation of posterior distributions is typically costly. Conversely, the computation of deterministic segmentation results, as for example by the popular active-contour approaches, is inexpensive and has shown to be an effective approach. Hence, we were motivated to merge both technologies, to obtain an active-contour inspired segmentation approach capable of estimating posterior distributions efficiently.

In previous work [50], we described an *Active Mean Fields* (AMF) approach to image segmentation that used a variational mean field method (VMF) approach along with a logit representation to construct a segmentation system similar to the one described here. This method empirically generated accurate segmentations and converged well, but used a different, and more awkward, approximation of the expected value of the length functional. In this present work, we use a linearization approach via the Plefka approximation. Using this approximation has profound consequences as it allows to make connections to the Chan-Vese [15] segmentation model and the ROF denoising model [51] in the continuous space. This connection in turn makes possible the efficient implementation of the AMF model through approaches used for ROF denoising. Hence, the overall model is convex, easy to implement and fast. Furthermore, we show good approximation properties in comparison to the "exact" distribution.

1.2. Contributions. The main contributions of this article are:

- It derives an AMF approach that allows a computationally efficient estimation of the posterior distribution of the segmentation label map based on the VMF approximation for binary image segmentation regularized via a boundary length prior.
- It establishes strong connections between the proposed AMF model, active-contour models and total-variation (TV) denoising. In particular, the model retains the global optimality of convex active-contours while estimating a level-set function that has a direct interpretation as an approximate posterior on the segmentation. This is in contrast to level-set techniques which use the

zero level-set only as a means for representing the object boundary with no (probabilistic) interpretation of the non-zero level-sets.

- It demonstrates how the Rudin-Osher-Fatemi (ROF) TV denoising model can be used to efficiently compute solutions of the AMF model. Hence, given the widespread availability of high-performance ROF-solvers, the AMF model is very simple to implement and will be immediately usable by the community with little effort.

1.3. Background. The earliest and simplest probabilistic image segmentation approaches frequently used pixel-wise independent Maximum Likelihood (ML) or MAP classifiers [56], that could be as simple as image thresholding. Better performance, in the face of noise, motivated the use of regularization, or prior probability models on the label fields that discouraged fragmentation [4], leading to the wide-spread application of Markov Random Field (MRF) models [28, 59]. Image segmentation with MRF models was initially thought to be computationally complex, which motivated approximations, including the *mean field* approach from statistical physics [31, 17]. Moreover, recently, fast solvers have appeared using graph-cuts, belief propagation or linear programming techniques that yield globally optimal solutions for certain energy functions [54].

Typically, the segmentation problem is posed as the minimization of an energy or negative log-likelihood that incorporates an image likelihood and a regularization term on the boundaries of segmented objects. This regularization may be specified either: (i) directly on the boundary (explicitly as a parametric curve or surface, or implicitly through the use of level-set functions); or (ii) by representing objects via indicator functions, where discontinuities in those functions identify boundaries. The direct boundary representation is attractive because it reduces complexity as only objects of co-dimension one need to be dealt with (*i.e.*, a curve in 2D, a surface in 3D, etc.). The price for this reduction in complexity is that, usually, minimization becomes non-convex, and hence can get trapped in poor local minima in the absence of good initializations. In the *snakes* approach [32], a popular example of explicit boundary representation, the boundary curve represented by control points is evolved such that it captures the object of interest (for example, by getting attracted to edges) while assuring regularity of the boundary by penalizing rapid boundary changes through elasticity and rigidity terms. Although computationally efficient, explicit parametric representations cannot easily deal with topological changes and have numerical issues due to their fixed object parameterization (*e.g.* when an initial curve grows or shrinks drastically). Furthermore, though not an intrinsic problem of explicit parameterizations, such methods are typically not geometric, making evolutions dependent on curve parameterizations.

In contrast, level-set representations [42, 36] of active-contour methods [10, 33] do not suffer from these topological and parameterization issues. These methods use implicit representations of the label-field, where an object's boundary is, for example represented through the zero level-set of a function. A parametric boundary representation is evolved *directly*, for example by moving its associated control points. For a level-set representation the *level-set function* is evolved, which *indirectly* implies an evolution of the segmentation boundary. Specifically, an evolution equation is imposed on the level-set function such that its zero level-set moves as desired. As the level-set function is (by construction) either strictly positive or negative (depending on convention used) inside the object and strictly negative or positive on the outside of the object, a labeling can be obtained by simple thresholding. Level-set approaches for

image segmentation make use of advanced numerical methods to solve the associated partial differential equations [42, 53]. To assure boundary regularity, segmentation energies typically penalize boundary length or surface area.

While initial curve and surface evolution methods focused on energy minimization based on boundary regularity and boundary misfit, later approaches such as the Chan-Vese model [15], added terms that encoded statistics about the regions inside and outside the segmentation boundary. Such region-based models can be as simple as homoscedastic (*i.e.*, same variance) Gaussian likelihoods with specified (but distinct) means for foreground and background respectively, as in the Chan-Vese case. They can also be much more complex such as trying to maximally separate intensity or feature distributions inside and outside an object [26]. Overall, a large variety of region-based approaches exist, providing great modeling freedom [20]. While region-based models are less sensitive to initialization, they are still non-convex when combined with weighted curve-length terms for regularization. Hence, a global optimum cannot be guaranteed by numerical optimization for such formulations. The dependency on curve and surface initializations popularized the formulation of energy minimization methods which can find a global energy optimum. One such approach is to reformulate an energy minimization problem as a problem defined over an appropriately chosen graph.

In the context of image segmentation, the idea is to create a graph with added source and sink nodes in such a way that a minimum cut of the graph implies a variable configuration which minimizes the original image segmentation energy [7]. For a large class of binary segmentation problems, these graph-cut approaches allow for the efficient computation of *globally* optimal solutions through max-flow algorithms [34]. In particular, discrete versions of the active-contour and Chan-Vese models (with fixed means) can be formulated. To avoid computing trivial solutions for the boundary-only active contours, graph-cut formulations add seed-points, specifying fixed background and foreground pixels or voxels (in 3D). While conceptually attractive, graph-cut approaches suffer from the need to build the graphs and the necessity to specify discrete neighborhood structures which may negatively affect the regularity of the obtained solution by creating so-called metrication artifacts.

Recently, the focus has shifted away from level-set and graph representations to formulations of active contours and related models by means of indicator functions [2, 8] defined in the continuum and allowing for convex formulations. These methods are closely related to segmentation via graph-cuts, but avoid the construction of graphs and can alleviate metrication artifacts. A key insight here is that the boundary-length or area term can be formulated through the total variation (TV) of an indicator function. This regularization formulation becomes convex when followed by relaxation of the indicator function to the interval $[0, 1]$. Hence these approaches strike an attractive balance between Partial Differential Equation (PDE)-based level-set formulations and the global properties of graph-cut methods. As they are specified via PDEs, highly accurate and fast implementations for these solvers are available [47]. As these convex formulations make use of TV terms, they are conceptually related to TV image-denoising. The use of TV regularization for denoising was pioneered by Rudin, Osher and Fatemi (the ROF model [51]). The ROF model uses quadratic (*i.e.*, ℓ_2) coupling to the image intensities and TV for edge-preserving noise-removal [9]. Approaches with ℓ_1 coupling yielding a form of geometric scale-space have also been proposed [13]. As we will see, our proposed approach will be closely related to these modern TV regularization and denoising approaches.

Segmentation approaches based on energy-optimization as discussed above typically either have a probabilistic interpretation (as negative log-likelihoods) or have been explicitly derived from probabilistic considerations. The reader is referred to Cremers et al. [20] for a review of recent developments in probabilistic level-set segmentation. All these techniques, while probabilistic in nature, seek optimal labels and do not directly provide information about the posterior distribution or uncertainty in their solutions. In contrast, our proposed AMF approach will approximate posterior distributions from which one can infer a segmentation and corresponding uncertainty.

1.4. AMF Segmentation Approach. AMF segmentation is a Bayesian approach, which results in a posterior distribution on the label map. The AMF approach combines explicit representations of label likelihoods with a boundary length prior. As we will show, our approach makes strong connections to ROF-denoising, and convex active-contour as well as probabilistic active-contour formulations.

In prior work, Monte-Carlo approaches have been used to characterize posterior distributions on segmentations, which require sampling [24, 18, 45]. In addition, the Monte-Carlo approach is quite general about statistical modeling assumptions so that it could be applied to the likelihood and regularity terms of our segmentation tasks. Approximations are then only caused by the sampling. A potential drawback of Monte-Carlo approaches is that an accurate estimation might require the generation of a large number of samples, which can be time consuming.

In contrast to the Monte-Carlo approach, our mean-field approximation is based on a factorized distribution that is quick to compute, but which is a relatively severe approximation. A potential drawback of our method is that samples drawn from the approximated posterior can have an un-natural fragmented appearance. However, our experimental results reveal that the approximation is surprisingly accurate (in terms of correlation of the posterior probabilities and the segmentation area), when compared to the exact model using much slower Gibbs sampling.

In summary, the primary advantages of our approach are speed, simplicity, and leverage of existing convex solver technology. We show in Section 3.2 that using ROF-denoising on the logit field of label likelihoods results in a "denoised" logit transform from which a label probability function can easily be obtained through a sigmoid transformation. Given an ROF solver, the AMF model can thus be implemented in one line of source-code. Furthermore, the AMF model provides a good approximation of the posterior of the segmentation under a curve-length prior as we experimentally show in Section 5.3.

1.5. Structure of the Article. In Section 2, we specify a discrete-space probabilistic formulation of segmentation with the goal of finding the posterior distribution of labels, given an input image. We use the VMF approach, along with a linearization approximation that simplifies the problem. This results in an optimization problem for determining the parameters of an approximation to the posterior distribution on pixel labels. In the style of Chan and Vese [15] and many others, we shift from discrete to continuous space facilitating use of the calculus of variations for the optimization problem, yielding the Euler-Lagrange equations for the AMF model.

In Section 3, we show that the AMF Euler-Lagrange equations for the zero level-set correspond to those of a special case of the Chan-Vese model [15], and that the AMF "approximate posterior" has the same mode, or MAP realization, as the exact

posterior distribution. Subsequently we show that the AMF Euler-Lagrange equations have the same form as those of the ROF model of image denoising, and we discuss methods that may be used for solving AMF.

Section 4 describes other important properties of AMF. We show that for a one-parameter family of realizations, the approximated and exact posteriors agree ratiometrically, and we explore their agreement for more general realizations. In addition, we show that the AMF problem is convex, and is unbiased in a particular sense.

Section 5 shows the experimental results on examples that include intensity ambiguities. It also demonstrates the quality of the AMF in approximating the true posterior via Gibbs sampling. Furthermore, Section 5 discusses AMF results for real ultrasound images of the heart, the prostate, a common test image in computer vision, and on a large collection of images from the `icgbench` segmentation dataset [52].

Finally, Section 6 concludes with a summary and an outlook on future work. Detailed derivations of the approximation properties can be found in the appendix.

2. Active Mean Fields (AMF). This section introduces the basic discrete-space probabilistic model (Section 2.1), that includes a conventional conditionally independent likelihood term and a prior that penalizes the boundary length of the labeling. The VMF approach is used (Section 2.2), along with a Plefka approximation (Section 2.3), to construct a factorized distribution that, given image data, approximates the posterior distribution on labelings. The resulting optimization problem for determining the parameters of the variational distribution has a KL-divergence data attachment term and a TV regularizer. The objective function is converted to continuous space (Section 2.4), yielding the Euler-Lagrange equations of the AMF model (Section 2.5), that involve logit label probabilities and likelihoods along with a curvature term.

In the following sections, we use upper-case P and Q to represent probability mass functions and lower-case p and q to represent probability density functions.

2.1. Original Probability Model. Define the space of images as a compact domain [1] $\mathcal{X} \subset \mathbb{R}^2$ indexed by $x \in \mathbb{R}^2$ and let $\mathbb{I}_{\mathcal{X}} \triangleq \{i : x_i \in \mathcal{X}\}$ denote the indices of the lattice of image pixels. Furthermore, Z denotes a binary random field defined on the pixel lattice whose realizations z are the binary labelings of a real-valued image y on the pixel lattice; given the image pixel index $i \in \mathbb{I}_{\mathcal{X}}$, z_i and y_i are the corresponding quantities specific to pixel $x_i \in \mathcal{X}$. For convenience, we write $p(\mathrm{y}_i|h) \triangleq p(\mathrm{y}_i|z_i = h)$ with $h \in \{0, 1\}$, where the definition of $p(\mathrm{y}_i|\mathrm{z}_i)$ is problem specific and is assumed to be given (for example, specified parametrically or obtained through kernel density estimation on a given feature space; we will not address this issue here). Now, if we make the usual assumption that the likelihood term, *i.e.*, the probability density of observing intensities conditioned on labels, is conditionally independent and identically distributed (iid), *i.e.*

$$p(\mathrm{y}|\mathrm{z}) = \prod_{i \in \mathbb{I}_{\mathcal{X}}} p\left(\mathrm{y}_i \mid \mathrm{z}_i\right), \tag{2.1}$$

[1] Our theory also holds for higher dimensions, *i.e.*, $\mathcal{X} \subset \mathbb{R}^n$. We discuss our approach in $\mathbb{R}^2$ for simplicity and hence talk about pixels. In 3D for example, we would deal with voxel grids and we would need to compute a 3D variant of total variation, but the overall results will hold unchanged.

then the corresponding log-likelihood, defined with respect to the logit transform of the pixel-specific image likelihood

$$\psi_i \triangleq \ln \frac{p(\mathrm{y}_i|1)}{p(\mathrm{y}_i|0)} \ , \tag{2.2}$$

is

$$\ln p(\mathrm{y}|\mathrm{z}) = \sum_{i \in \mathbb{I}_{\mathcal{X}}} \ln p\left(\mathrm{y}_i \mid \mathrm{z}_i\right) = \sum_{i \in \mathbb{I}_{\mathcal{X}}} \mathrm{z}_i \psi_i + \ln p(\mathrm{y}_i|0). \tag{2.3}$$

Next, we apply a prior that penalizes the length $L(z)$ of the boundaries of the label map,

$$P(\mathrm{z}) \propto \exp(-\lambda L(\mathrm{z})). \tag{2.4}$$

Here, $\lambda \in \mathbb{R}^+$ is a constant. The larger λ the more irregular segmentation boundaries are penalized and therefore discouraged. We defer discussion of the length functional $L(\cdot)$ to Section 2.4.

By Bayes' rule the posterior probability of the label map given the image is

$$P(\mathrm{z}|y) \propto \ p(y|\mathrm{z})P(\mathrm{z}) \tag{2.5}$$

so that

$$\ln P(\mathrm{z}|y) = \sum_{i \in \mathcal{X}} \left[\mathrm{z}_i \psi_i + \ln p(\mathrm{y}_i|0)\ \right] - \lambda L(\mathrm{z}) + \mathrm{const}\,. \tag{2.6}$$

Here the constant is equal to the log-partition function of the prior distribution. This constant is not easily computed, as it requires a sum over all of the configurations of z.

2.2. Variational Mean-Field Approximation. As mentioned above, the partition function cannot easily be computed. In the variational mean-field (VMF) approach [58], we approximate the posterior distribution P via a simpler variational distribution Q by minimizing the distance between P and Q (here, in a Kullback-Leibler sense – see details below). The explicit computation of the integrals involved in the partition function (for continuous variables) can thereby be avoided. Specifically, the mean-field method approximates the joint distribution of a countable family of random-variables as a product of univariate distributions. The VMF approximation is widely used in machine learning and other fields [58].

For the binary segmentation problem, we define the mean-field approximation $Q(\mathrm{z};\theta)$ of the posterior distribution $P(\mathrm{z}|y)$ as a field of independent Bernoulli random variables z_i defined on the lattice $i \in \mathbb{I}_{\mathcal{X}}$ with probability θ_i, which constitute the random field Z:

$$Q(\mathrm{z};\ \theta) \triangleq \prod_{i \in \mathbb{I}_{\mathcal{X}}} \theta_i^{\mathrm{z}_i} \left(1-\theta_i\right)^{1-\mathrm{z}_i} \tag{2.7}$$

$$= \exp\left\{ \sum_{i \in \mathbb{I}_{\mathcal{X}}} \left[\mathrm{z}_i \phi_i + \ln\left(1-\theta_i\right)\right] \right\}, \tag{2.8}$$

where $\phi_i \triangleq \ln \frac{\theta_i}{1-\theta_i}$. Next, the parameters of $Q(\mathrm{z};\cdot)$ are set so that it minimizes the KL-divergence with respect to the original probability model, *i.e.*,

$$\hat{\theta}_* \triangleq \arg\min_{\theta} \mathbb{KL}[Q(\mathrm{z};\theta)||P(\mathrm{z}|y)] \tag{2.9}$$

$$= \arg\min_{\theta} \mathbb{E}_Q\left[\ln Q(\mathrm{z};\theta) - \ln P(\mathrm{z}|y)\right] \tag{2.10}$$

$$= \arg\min_{\theta} \mathbb{E}_Q\left[\ln Q(\mathrm{z};\theta) - \ln p(y|\mathrm{z}) + \lambda L(\mathrm{z})\right]. \tag{2.11}$$

With minor abuse of KL-divergence notation:

$$\hat{\theta}_* = \arg\min_{\theta} \mathbb{KL}[Q(\mathrm{z};\theta)||p(y|\mathrm{z})] + \mathbb{E}_Q\left[\lambda L(\mathrm{z})\right]. \tag{2.12}$$

In other words, the VMF approximation selects the parameters of the factorized variational distribution $Q(\mathrm{Z};\ \theta)$ such that (i) local image likelihood information, $p(y|\mathrm{z})$, is captured *while at the same time* (ii) considering the expected value of the segmentation boundary length (which is a global property that regularizes the solution).

2.3. Plefka's Approximation. Although minimizing the KL-divergence term in Eqn. (2.12) with respect to θ is relatively straightforward, minimizing $\mathbb{E}_Q\left[L(\mathrm{z})\right]$ is generally not as it entails a sum over all configurations of z. In the mean-field literature, difficult expectations of functions of random-fields have been approximated using Plefka's method [46].

Noting that $\mathbb{E}_Q\left[\mathrm{z}\right] = \theta$ according to Eqn. (2.7) and that the first order Taylor expansion of the curve length function about z^* is $L(\mathrm{z}) \approx L(\mathrm{z}^*) + (\mathrm{z} - \mathrm{z}^*) \cdot \nabla L(\mathrm{z}^*)$, then Plefka's approximation states that

$$\mathbb{E}_Q\left[L(\mathrm{z})\right] \approx L(\mathrm{z}^*) + (\mathbb{E}_Q\left[\mathrm{z}\right] - \mathrm{z}^*)\nabla L(\mathrm{z}^*) \approx L(\mathbb{E}_Q\left[\mathrm{z}\right]) = L(\theta) \tag{2.13}$$

so that an approximation of Eqn. (2.12) is

$$\hat{\theta} \triangleq \arg\min_{\theta} \mathbb{KL}[Q(\mathrm{z};\theta)||p(y|\mathrm{z})] + \lambda L(\theta), \tag{2.14}$$

where $\hat{\theta}_* \approx \hat{\theta}$.

Assuming $L(\cdot)$ is convex, as in the present case, then the Plefka approximation of Eqn. (2.14) is a lower bound to the original objective function of Eqn. (2.12) as Jensen's inequality states $\mathbb{E}_Q\left[L(\mathrm{z})\right] \geq L(\mathbb{E}_Q\left[z\right]) = L(\theta)$. While this is not directly useful for our purposes, there has been some work on "converse Jensen inequalities" [23] that may provide useful bound relationships. In the end, approximations are justified by the quality of their results, such as the favorable properties highlighted in Section 4.

2.4. Continuous Variant of Variational Problem. In the previous section, we showed how the problem of computing the posterior distribution of a label-field under an (unspecified) boundary-length prior results in solving the optimization problem of Eqn. (2.14). To solve this problem using computationally efficient PDE optimization techniques, we first replace the random-field defined on a discrete lattice by one defined on continuous space.

Expanding Eqn. (2.14) by using the definition of the log likelihood (Eqn. (2.3)) and of $Q(\cdot,\cdot)$ (Eqn. (2.8)) we get:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_Q \left[\sum_{i \in \mathbb{I}_{\mathcal{X}}} \mathrm{z}_i \phi_i + \ln(1-\theta_i) - \mathrm{z}_i \psi_i - \ln p(\mathrm{y}_i|0) \right] \tag{2.15}$$

$$+ \lambda L(\theta) \tag{2.16}$$

$$= \underset{\theta}{\operatorname{argmin}} \sum_{i \in \mathbb{I}_{\mathcal{X}}} [\theta_i \phi_i + \ln(1-\theta_i) - \theta_i \psi_i - \ln p(\mathrm{y}_i|0)] + \lambda L(\theta). \tag{2.17}$$

$$= \underset{\theta}{\operatorname{argmin}} \sum_{i \in \mathbb{I}_{\mathcal{X}}} [\theta_i \phi_i + \ln(1-\theta_i) - \theta_i \psi_i] + \lambda L(\theta). \tag{2.18}$$

To solve the above equation by extending θ to the continuum, the logit transform of the likelihood is now defined as

$$\psi(x) \triangleq \ln \frac{p(\mathrm{y}(x)|\mathrm{z}(x)=1)}{p(\mathrm{y}(x)|\mathrm{z}(x)=0)}, \tag{2.19}$$

where x denotes the location (*i.e.*, the continuous equivalent of the index $i \in \mathbb{I}_{\mathcal{X}}$), and $\mathrm{y}(x)$, $\mathrm{z}(x)$, and $\theta(x)$ are the corresponding values of y, z, and θ at location x. Similarly, the continuous variant of the logit transform of the variational probability function, $\theta(x)$, is defined as

$$\phi(x) \triangleq \ln \frac{\theta(x)}{1-\theta(x)}. \tag{2.20}$$

Now, if we denote with v the area of a lattice element and replace the summation over the lattice with integration over $\mathcal{X}$, then Eqn. (2.18) becomes in the continuous space,

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}}\ v^{-1} \cdot \left(\int_{\mathcal{X}} \theta(x)(\phi(x) - \psi(x)) + \ln(1-\theta(x))\, dx \right) + \lambda L(\theta)\ . \tag{2.21}$$

By the co-area formula [5], the length of the boundaries of a binary image defined on the continuum is equal to its total-variation:

$$L(\mathrm{z}) = \mathrm{TV}[\mathrm{z}(x)] = \int_{\mathcal{X}} ||\nabla \mathrm{z}(x)||_2\ dx \tag{2.22}$$

where $||\cdot||_2$ is the 2-norm and $\nabla \mathrm{z}$ is the (weak) gradient of z.[2]

Therefore putting it all together, the continuous variant of the variational problem is:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \int_{\mathcal{X}} \theta(x)(\phi(x) - \psi(x)) + \ln(1-\theta(x)) + v\lambda ||\nabla \theta(x)||_2\ dx, \tag{2.23}$$

which we call the *Active Mean Field approximation.* Note, that $\phi(x)$ depends on $\theta(x)$ according to Eqn. (2.20).

[2] $\nabla(z)$ is defined as $\int_{\mathcal{X}} \langle \nabla \mathrm{z}, \eta \rangle dx = -\int_{\mathcal{X}} \langle \mathrm{z}, \nabla \cdot \eta \rangle\ dx$ for any test function $\eta : \mathcal{X} \to \mathbb{R}^2$; in the case of $\mathrm{z}(x)$ being an element of a convex set, $L(z)$ is convex.

2.5. Euler-Lagrange (EL) Equations. Defining the curvature operator,

$$\kappa(\theta(x)) \triangleq \nabla \cdot \left(\frac{\nabla\theta(x)}{\|\nabla\theta(x)\|_2} \right), \tag{2.24}$$

the Euler-Lagrange equation describing the stationary points of Eqn. (2.23) is given by:

$$\phi(x) - \psi(x) - v\lambda\kappa(\theta(x)) = 0 \ . \tag{2.25}$$

This can be derived as follows: Expanding $\phi(x)$ according to Eqn. (2.20), we obtain the objective function

$$E(\theta) = \int_{\mathcal{X}} -\theta(x)\psi(x) + \theta(x)\ln(\theta(x)) + (1-\theta(x))\ln(1-\theta(x)) + v\lambda\|\nabla\theta(x)\|_2 \, dx. \tag{2.26}$$

The variation of $E(\theta)$ is [55]

$$\delta E(\theta; \delta\theta) \triangleq \frac{\partial E(\theta + \epsilon\delta\theta)}{\partial\epsilon}|_{\epsilon=0} \ , \tag{2.27}$$

where $\epsilon \in \mathbb{R}$, $\delta\theta$ denotes an admissible perturbation of $E(\theta)$, and $\frac{\partial}{\partial\epsilon}$ denotes the partial derivative with respect to ϵ. The variation becomes

$$\delta E(\theta; \delta\theta) = \int_{\mathcal{X}} -\psi(x)\delta\theta(x) + \ln\frac{\theta(x)}{1-\theta(x)}\delta\theta(x) + v\lambda\frac{1}{\|\nabla\theta(x)\|_2}\nabla\theta(x)\cdot\nabla\delta\theta(x) \, dx. \tag{2.28}$$

Integration by parts assuming Neumann boundary conditions and using Eqn. (2.20) results in

$$\delta E(\theta; \delta\theta) = \int_{\mathcal{X}} \left(\phi(x) - \psi(x) - v\lambda\nabla\cdot\left(\frac{\nabla\theta(x)}{\|\nabla\theta(x)\|_2}\right)\right)\delta\theta(x) \, dx \ . \tag{2.29}$$

As the variation needs to vanish for all admissible perturbations $\delta\theta(x)$ at optimality, we obtain the Euler-Lagrange equation

$$\phi(x) - \psi(x) - v\lambda\kappa(\theta(x)) = 0 \ . \tag{2.30}$$

According to Eqn. (2.20), $\phi(x)$ is obtained from $\theta(x)$ through a logit transform. Consequentially, we can obtain $\theta(x)$ from $\phi(x)$ via the sigmoid function

$$\sigma(x) \triangleq (1 + \exp(-x))^{-1} \tag{2.31}$$

as $\theta(x) = \sigma(\phi(x))$. The sigmoid function, $\sigma(\cdot)$, is monotonic (*i.e.*, $\sigma'(x) > 0$) so that

$$\nabla\theta(x) = \sigma'(\phi(x))\nabla\phi(x) \tag{2.32}$$

and

$$\frac{\nabla\theta(x)}{\|\nabla\theta(x)\|_2} = \frac{\sigma'(\phi(x))\nabla\phi(x)}{\|\sigma'(\phi(x))\nabla\phi(x)\|_2} = \frac{\nabla\phi(x)}{\|\nabla\phi(x)\|_2} \ . \tag{2.33}$$

Hence, the Euler-Lagrange equation can be rewritten as

$$\phi(x) - \psi(x) - v\lambda\kappa(\phi(x)) = 0 \ . \tag{2.34}$$

In summary, the distribution $Q(\mathrm{z};\theta)$ approximates the "exact" distribution, $P(\mathrm{z}|y)$, in the KL-divergence sense when ϕ (the logit transform of the parameter θ) satisfies the Euler-Lagrange equation of the AMF model; we will refer to Eqn. (2.34) as the "AMF Equation." As the objective function is strictly convex (see Section 4.2) in θ, the stationary point is the unique global optimum.

3. Connections to Chan-Vese and ROF. In this section we establish the connection between the AMF model and the Chan-Vese segmentation model (Section 3.1) as well as the ROF denoising model (Section 3.2). In particular, we show that the Chan Vese Euler-Lagrange equations correspond to those of the zero level-set of the AMF model, so a Chan-Vese segmentation can be obtained as the zero level-set of the AMF solution. We also show that the AMF Euler-Lagrange equations (Eqn. (2.34)) have the same form as those of the ROF model. Therefore, the solver technologies that have been developed for the ROF model may be deployed for AMF.

3.1. Connection to Chan-Vese. To derive the connection between the AMF and the Chan-Vese approach, we introduce the energy $E_{cv}(\cdot)$ for the generalized Chan-Vese model based on a relaxed indicator function (*i.e.*, $\theta \in [0,1]$), which, according to [8], can be written as

$$E_{cv}(\theta) = \int_{\mathcal{X}} -\theta(x)\psi(x) + v\lambda\|\nabla\theta(x)\|_2 \, dx, \tag{3.1}$$

with the first part of the function being the data term and the second term regularizing the boundary length. Such a length prior is essential to encourage large, contiguous segmentation areas. The importance of the length-prior becomes especially clear in the context of the Mumford-Shah model [39], of which the Chan-Vese model is a special case. In the absence of a length prior, the Mumford-Shah approach will assign each pixel in regions with constant image intensity to its own (separate) parcel. The standard Chan-Vese model [16] (without the area prior of this model) can be recovered from Eqn. (3.1) for the special case that the class conditional intensity model is Gaussian, *i.e.*, $y_i|\mathrm{z}_i = 1 \sim N(\mu_1, \sigma_1^2)$ and $y_i|\mathrm{z}_i = 0 \sim N(\mu_0, \sigma_0^2)$. In this case:

$$\psi(x) \triangleq \ln\frac{\sigma_0}{\sigma_1} - \frac{1}{2\sigma_0^2}(y(x)-\mu_0)^2 + \frac{1}{2\sigma_1^2}(y(x)-\mu_1)^2, \tag{3.2}$$

and the corresponding Chan-Vese energy becomes:

$$E_{cv}^{gauss}(\theta) = \int_{\mathcal{X}} -\theta(x)\left(\ln\frac{\sigma_0}{\sigma_1} - \frac{1}{2\sigma_0^2}(y(x)-\mu_0)^2 + \frac{1}{2\sigma_1^2}(y(x)-\mu_1)^2\right) + v\lambda\|\nabla\theta(x)\|_2 \, dx. \tag{3.3}$$

The means of the Gaussians (μ_1, μ_2) are estimated jointly in the standard Chan-Vese model [15] and the standard deviations are assumed to be fixed and identical. In contrast, in the generalized Chan-Vese model (Eqn. (3.1)), parameters of $\psi(x)$ are typically assumed to be fixed and are not jointly estimated. This assures the convexity of the overall model. However, if desired, these parameters can also be estimated. A simple approach would be an alternating optimization strategy. Note that the Chan-Vese segmentation model of Eqn. (3.3) becomes Otsu-thresholding [43] if the length prior is disregarded ($\lambda = 0$). Hence, unlike Chan-Vese segmentation, Otsu-thresholding cannot suppress image fragmentation and irregularity.

The Euler-Lagrange equations of the generalized Chan-Vese energy (Eqn. (3.1)) are:

$$-\psi(x) - v\lambda\kappa(\theta(x)) = 0. \tag{3.4}$$

This is identical to the AMF Euler-Lagrange equation (Eqn. (2.30)) at the zero level-set $\phi(x) = 0$. By construction, the zero level-set of a *level-set implementation for the*

generalized Chan-Vese model has to agree with the solution obtained from the *Euler-Lagrange equations of the generalized Chan-Vese model using indicator functions* as both minimize the same energy function just using different parameterizations. Consequentially, also, the zero level-sets of both the AMF model and the level-set implementation of Chan-Vese need to agree.

In contrast to the generalized Chan-Vese model described above, the original Chan-Vese model of [15], formulated as a curve evolution approach, is characterized by an energy function (penalizing segmentations with large, continuous areas) with an additional term of the form

$$E_{area}(C) = \nu \text{Area}(inside(C)), \tag{3.5}$$

where C denotes the curve defining the boundary of the segmentation, $\nu \in \mathbb{R}_+^0$ is a non-negative constant to weight the area influence, and Area($inside(C)$) simply denotes the area enclosed by C. For implementation purposes C is implicitly represented by the zero level-set of a *level set function* ϕ. The corresponding Euler-Lagrange equation is, on the zero level-set [15],

$$-\psi(x) - v\lambda\kappa(\phi(x)) + \nu = 0 \ . \tag{3.6}$$

Examining the ν level-set of the AMF model (Eqn. (2.34)), so that $\phi(x) = \nu$, we notice that this level-set satisfies the same Euler-Lagrange equation as the zero level-set of the Chan-Vese model with a specified non-zero value of ν. In other words, the level-sets of the dense AMF solution provide a family of solutions for the Chan-Vese problem for a continuum of values of the area penalty.

Note that such area penalties cannot effectively be added in the indicator-function based approaches to the Chan-Vese active-contour models proposed by Appleton et al. [2] and Bresson et al. [8]. The goal of these models is to capture a binary segmentation result through a relaxed indicator function, (*i.e.*, $\theta \in [0,1]$ instead of $\theta \in \{0,1\}$). However, it can be shown [41] that in certain instances this relaxation produces undesirable segmentation results when combined with an area penalty.

3.2. Connection between AMF and ROF Models. In their seminal paper, Rudin, Osher and Fatemi [51] proposed a denoising method for, e.g., intensity images $u_0(x)$,

$$u^*(x) = \arg\min_u \int_{\mathcal{X}} \|\nabla u(x)\|_2 \ dx \quad \text{s.t.} \quad \int_{\mathcal{X}} (u(x) - u_0(x))^2 \ dx = \sigma^2 \ , \tag{3.7}$$

where $\sigma > 0$. As discussed by Vogel and Oman [57], this is equivalent to the following unconstrained problem,

$$u^*(x) = \arg\min_u \left[\int_{\mathcal{X}} (u(x) - u_0(x))^2 dx \ + \frac{\alpha}{2} \int_{\mathcal{X}} \|\nabla u(x)\|_2 dx \right] \ , \tag{3.8}$$

for a suitable choice of $\alpha > 0$. They refer to this formulation as "TV penalized least squares."

The corresponding Euler-Lagrange equation is

$$u(x) - u_0(x) - \alpha\kappa(u(x)) = 0 \ . \tag{3.9}$$

For $\alpha = v\lambda$, this equation has the same form as the Euler-Lagrange equations of the AMF model of Eqn. (2.34), which is

$$\phi(x) - \psi(x) - v\lambda\kappa(\phi(x)) = 0 \ . \tag{3.10}$$

In this equivalence, the denoised intensity image of the ROF model, u, corresponds to the logit parameter field of the AMF distribution, ϕ, while the noisy input intensity image of the ROF model, u_0, corresponds to the logit-transformed label probabilities in the AMF problem, ψ. Furthermore, if the class conditional intensity model is homoscedastic Gaussian, then (from Eqn. (3.2)) $\psi(x)$ is linear in the observed intensity. Furthermore, the AMF solution is equivalent to solving an ROF problem that is effectively denoising the logit-transformed label likelihoods.

Because of the equivalence of the Euler-Lagrange equations of the AMF and the ROF models, the considerable technology developed for solving the ROF model may be applied to the AMF model. In particular, a globally optimal solution (see Section 4.2 for a proof of the convexity of this model) of the AMF model can be computed by the ROF denoising approach. In other words, given an ROF solver (ROFsolve) that minimizes

$$E_{ROF}(u; u_0, \alpha) = \int_{\mathcal{X}} (u(x) - u_0(x))^2 + \frac{\alpha}{2} \|\nabla u\|_2 \, dx \tag{3.11}$$

such that

$$u^* \triangleq \arg\min_u E_{ROF}(u; u_0, \alpha) = \text{ROFsolve}(u_0, \alpha), \tag{3.12}$$

solving the AMF problem for a given ψ and λ then simply becomes

$$\theta^* = \sigma(\text{ROFsolve}(\psi, v\lambda)). \tag{3.13}$$

Eqn. (3.13) is *the central result concerning the implementation of our method* as it connects the optimal AMF solution to a straightforward ROF denoising problem.

4. Additional Properties of AMF. We now summarize some approximation properties of AMF (Section 4.1), show the objective function to be convex (Section 4.2), and show that AMF is unbiased in a specific sense (Section 4.3).

4.1. Approximation Properties. Our goal is an efficient yet accurate approximation, $Q(\mathrm{z};\theta)$, to the exact posterior distribution $P(\mathrm{z}|y)$ for general realizations of z. To show that $Q(\mathrm{z};\theta)$ is in fact a good approximation, we study its properties here. For convenience, we only summarize the results of some of the approximation properties of the AMF model and refer to the appendix for mathematical details. In particular, the appendix shows that

a) The zero level-set of ϕ is the boundary of the most probable realization z_0 of $Q(\mathrm{z};\theta)$ *and* it defines the MAP realization under $P(\mathrm{z}|y)$. This is not generally the case for mean field approximations.
b) Because the log partition function of the prior is not easily computed we compare $\ln \frac{Q(\mathrm{z};\theta)}{Q(\mathrm{z}_0;\theta)}$ with $\ln \frac{P(\mathrm{z}|y)}{P(\mathrm{z}_0|y)}$, where z_0 is the most probable realization under both distributions according to a). These probability ratios are not only in agreement for the zero level-set, but also for realizations that are bounded by any level-set of ϕ.
c) The probability ratios approximately agree for realization whose boundary normals are close in direction to $\nabla\phi$.
d) If neither a) nor b) hold, the probability ratio for $Q(\mathrm{z},\theta)$ will be larger than that for $P(\mathrm{z}|y)$, *i.e.*, it underestimates the length penalty associated with the prior.

4.2. Convexity. A nice property of the AMF model is that its energy is strictly convex and therefore we can find a unique global minimizer. This is in contrast to the TV based segmentation models [2, 8] which are generally convex (but not strictly so) and therefore may have multiple non-unique optima.

To show convexity, we consider the continuum formulation of AMF which can be rewritten as a function of $\theta(x) \in [0, 1]$, as:

$$E_{\text{amf}}(\theta) = \int_{\mathcal{X}} -\theta\psi + v\lambda\|\nabla\theta\|_2 + (1-\theta)\ln(1-\theta) + \theta\ln\theta \; dx \tag{4.1}$$

where dependencies on space are dropped only for notational convenience (*i.e.*, $\theta = \theta(x)$ and $\psi = \psi(x)$) and we expressed ϕ in terms of θ. The term

$$\int_{\mathcal{X}} -\theta\psi + \lambda\|\nabla\theta\|_2 \; dx \tag{4.2}$$

is convex in θ as the first summand is linear in θ, the 2-norm is convex, ∇ is a linear operator and both terms are summed with a positive weight λ. To see that the rest of the integrand is also convex, consider a function of the form

$$f(\theta) = (1-\theta)\ln(1-\theta) + \theta\ln(\theta). \tag{4.3}$$

which implies that

$$f''(\theta) = \frac{1}{\theta(1-\theta)} > 0 \quad \text{for} \quad \theta \in (0, 1). \tag{4.4}$$

Therefore, $\int_{\mathcal{X}}(1-\theta)\ln(1-\theta) + \theta\ln(\theta)\; dx$ is strictly convex. Because the sum of convex and strictly convex functions is strictly convex, the overall AMF energy is strictly convex in θ and therefore has a unique global minimizer (see [6] for details on convexity preserving operations). In particular, we note that for a non-informative data term, *i.e.*, pixels are locally equally likely to be foreground or background ($\psi = 0$), $\theta(x) = \frac{1}{2}$ is the globally optimal solution. For the related standard TV segmentation model [2], which would only minimize Eqn. (4.2), any constant solution would be a global minimizer.

4.3. Unbiased in Homogeneous Regions. In this section we analyze the behavior of the AMF estimator over homogeneous (*i.e.*, constant intensity) patches of an image. The AMF objective function, Eqn. (2.23), can be written:

$$\hat{\theta} = \arg\min_{\theta} \int_{\mathcal{X}} \mathbb{KL}[Q(\mathrm{z}(x); \theta(x)) || p(y(x)|\mathrm{z}(x))]dx + v\lambda \mathrm{TV}[\mathrm{z}] \; . \tag{4.5}$$

Now, for a patch $\mathcal{X}$ of constant intensity, *i.e.*, $\psi(x) = \psi_0$, the optimum will be attained at $\ln(\theta(x)/(1-\theta(x))) = \phi(x) = \psi_0$ as both the KL and TV terms vanish. This in turn implies that the regularizer does not interact in homogeneous regions and an unbiased probability estimate is obtained.

In contrast, other probabilistic segmentation approaches, *e.g.* the Ising model [29], lack this "unbiased in homogeneous regions" property and because of this interaction with the regularizer, setting the regularization parameter λ in such cases can be tricky. To illustrate this point, consider a VMF treatment of the Ising model that parallels

the approach and notation used for AMF. Defining an Ising model where $N(i)$ are the neighbors of x_i and the neighborhood potential term is

$$U(\mathrm{z}) \triangleq \lambda \sum_{i \in \mathbb{I}_{\mathcal{X}}} \sum_{j \in N(i)} \mathrm{z}_i(1-\mathrm{z}_j) + (1-\mathrm{z}_i)\mathrm{z}_j, \tag{4.6}$$

then

$$P(\mathrm{z}|y) \propto p(y|\mathrm{z})P(\mathrm{z}) \text{ where } P(\mathrm{z}) \propto e^{-U(\mathrm{z})} \ . \tag{4.7}$$

Using the VMF approximation, we obtain:

$$\hat{\theta} = \arg\min_{\theta} \left\{ \mathbb{KL}[Q(\mathrm{Z};\theta)||p(y|\mathrm{z})] + \lambda \mathbb{E}_{Q(\mathrm{Z};\theta)}\left[U(\mathrm{Z})\right] \right\} \tag{4.8}$$

$$= \arg\min_{\theta} \left\{ \mathbb{KL}[Q(\mathrm{Z};\theta)||p(y|\mathrm{z})] + \lambda \sum_{i} \sum_{j \in N(i)} \theta_i(1-\theta_j) + (1-\theta_i)\theta_j \right\}, \tag{4.9}$$

which yields the following stationary-point equation:

$$\phi_i - \psi_i - 4n\lambda \sum_{j \in N(i)} \left[\theta_j - \frac{1}{2}\right] = 0, \quad n \triangleq |\{j \in N(i)\}|. \tag{4.10}$$

This consistency equation characterizes the solution of the VMF approximation to the Ising model. It is clear from Eqn. (4.10) that the regularization term will only be zero when the neighborhood average of θ_i equals $\frac{1}{2}$, while in other cases the unbiased property will not apply.

5. Experiments. This section illustrates the behavior of the proposed AMF model. Section 5.1 describes our numerical solution approach for the ROF model. Section 5.2 compares the AMF model to the standard Chan-Vese approach when dealing with ambiguous boundaries. Section 5.3 investigates how well the AMF model agrees with the original probability model without approximations. Section 5.4.1 qualitatively assesses the AMF model on real ultrasound data of the heart and the prostate, as well as on the Fabio image often used for testing in computer vision. Section 5.4.2 quantitatively analyzes AMF by applying it to the images from the `icgbench` segmentation benchmark dataset.

5.1. Numerical Solution. We indirectly solve the AMF model by relating it to the ROF problem as discussed in Section 3.2. The ROF model was initially solved [51] using a gradient descent method, and this may still be a reasonable option if AMF solving is embedded in an outer iteration, e.g. expectation-maximization [22]. The difficulty in computing the optimum of the ROF energy is due to the TV term that is not differentiable everywhere. The very first solver changed the optimization problem by replacing the TV term with $\sqrt{|\nabla u|^2 + \beta^2}$ [14], which made the energy function differentiable everywhere. To allow for better discretization of the TV term, primal-dual [14], and fully dual methods [11] have been explored. More recently, methods based on accelerated proximal gradient descent (`FISTA`) [3] and split Bregman iterations [27] have been applied to solve the ROF model. See [12] for a comprehensive overview of recent continuous optimization strategies for the ROF model. We use `FISTA` for all our following experiments on synthetic and real data. To avoid computational issues in our experiments, probabilities were clamped to be in $[10^{-5}, 1-10^{-5}]$. We used the Matlab `FISTA` implementation by Amir Beck and Marc Teboulle [3]. Convergence for `FISTA` was left at the default value of 10^{-4}. The maximum number of iteration steps was set to 10,000 but was never reached.

5.2. Segmentation with Ambiguity. A goal of AMF is to provide label probabilities from which the MAP solution for the segmentation can be obtained, but which also allow the assessment of segmentation uncertainty. To test this behavior, we generated a highly ambiguous segmentation scenario, depicted in Fig. 1. We start by assuming class conditional intensity distributions for the foreground and the background classes (Fig. 1 right). Specifically, the class conditional intensity distributions were obtained as a mixture of Gaussians. We use three Gaussians with means $\mu = \{30, 50, 70\}$ and corresponding standard deviations $\sigma = \{5, 10, 5\}$ and mix the first two ($\mu = \{30, 50\}$; $\sigma = \{5, 10\}$) to obtain the background conditional intensity distribution and the last two ($\mu = \{50, 70\}$; $\sigma = \{10, 5\}$) to obtain the foreground conditional intensity distribution. In both cases, the mixing coefficients are 0.5. The intensity distribution of the circle in the center of the image was chosen such that half of the circle has intensities that lie exactly in the middle between the foreground and background. In particular, the intensity of the region outside the circle is $\mu = 30$, the intensity of the upper part of the circle is $\mu = 50$, and the intensity of the lower part of the circle is $\mu = 70$. Gaussian noise with mean zero and standard deviation of $\sigma = 5$ was added to the background, $\sigma = 10$ to the upper part of the circle, and $\sigma = 5$ to the lower part of the circle. The results were obtained by assuming we know the conditional distributions for the foreground and background classes; likelihoods were computed based on the noisy data. The regularization term was weighted with $\lambda = 5$.

Fig. 2 (left two images) shows the local label probabilities for the noisy input image and for the noise-free image (that will not be available in practice). Fig. 2 (right two images) shows the label probabilities after running the AMF model (left) and after thresholding (binarization) at $P = 0.5$ (right) that also corresponds to the MAP solution. Note that neither the foreground probability is one nor the foreground probability is zero due to the chosen class conditional intensity distributions: both the means of the background ($\mu = 30$) and the foreground ($\mu = 70$) have non-zero likelihood for background *and* foreground. As desired, the AMF model captures the segmentation uncertainty by estimating the upper part of the circle at a probability close to $P = 0.5$. At the same time, due to spatial regularization, the AMF model removes noise effects. The MAP solution captures the most likely foreground area, but completely loses the ambiguous area.

Fig. 3 shows the estimated label probabilities and their true local counterparts along with a subtraction. The AMF method has effectively estimated the true label probabilities. Note that the true local label probabilities do not incorporate the effect of regularization. Hence, these two probabilities will slightly disagree at the segmentation boundaries.

5.3. Agreement with the Original Probability Model. In order to evaluate agreement between the original probability model, Eqn. (2.6), *i.e.*

$$\ln P(\mathrm{z}|y) = v^{-1} \sum_{i \in \mathcal{X}} [\mathrm{z}_i \psi_i + \ln p(\mathrm{y}_i|0)\ ] - \lambda L(\mathrm{z}) + \mathrm{const}\,. \tag{5.1}$$

and the AMF approximation, we conducted the following set of experiments on synthetic images. A binary random field was generated by sampling on a 100×100 grid from a Gaussian process with Matérn covariance function [21] with order parameter p and scale parameter l, that provides fine-grained control over the smoothness of the field. This continuous valued image was then thresholded at a quantile value selected uniformly at random to create the *ground truth binary label map* $\hat{\mathrm{z}}$ to which Gaussian

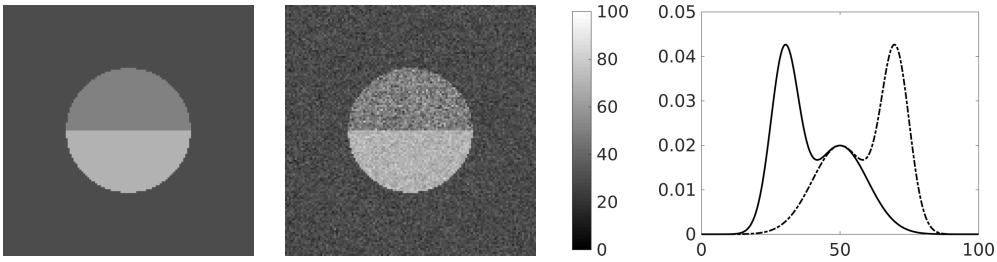


FIG. 1. *Ambiguous segmentation scenario. Left: original image, middle: noisy image, right: class conditional distributions. Distributions clearly overlap which should result in a segmentation ambiguity for the upper part of the circle which was deliberately chosen to have intensities in between the background and the foreground (bottom part of the circle). Background class conditional distribution displayed as a solid black line, foreground class conditional distribution displayed as a dash-dotted black line.*

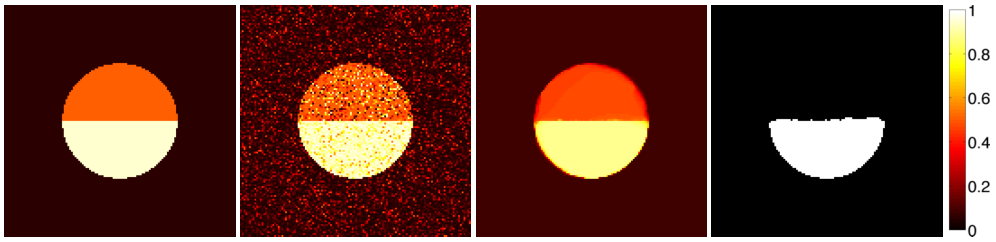


FIG. 2. *(a) Noise-free foreground label probabilities based on the noise-free image of Fig. 1 (which is not available in practice). (b) Noisy label probabilities based on the noisy image of Fig. 1. The upper part of the circle is clearly ambiguous with foreground label probability of $P = 0.5$. (c) Estimated label probabilities using the AMF model. (d) Estimated MAP solution (binarization at $P = 0.5$) from the AMF-estimated label probabilities. Clearly, the AMF model captures more information – the MAP solution completely loses the ambiguity of the upper part of the circle.*

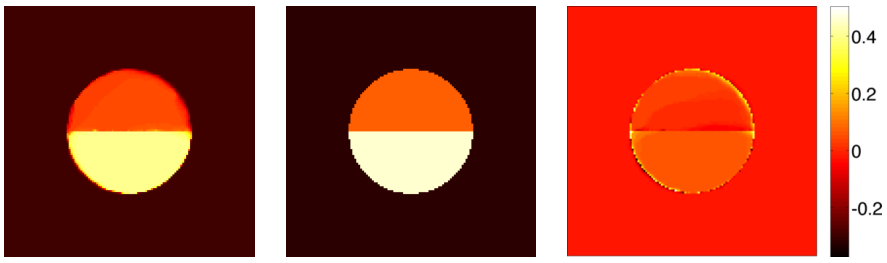
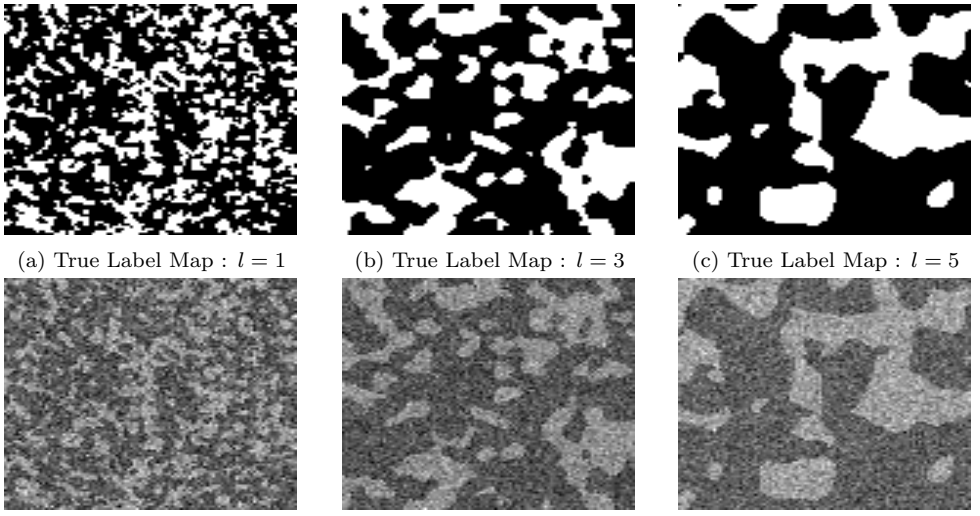


FIG. 3. *Left: estimated label probabilities by the AMF model. Middle: noise-free label probabilities. Right: difference between the probabilities. Differences exist primarily at the segmentation boundaries, which is expected since the AMF model includes spatial regularization effects while the noise-free label probabilities are computed strictly locally. Overall, there is a good agreement between the probabilities.*

noise is added to create a noisy image y. For our experiments, we set the order parameter $p = 1$ while varying the length scale parameter $l = 1, 3, 5$ and the noise standard deviation σ between 0.25 and 0.4. Increasing l produces label maps with smoother boundaries and larger contiguous regions. Single realizations of $\hat{z}$ for $l = 1, 3, 5$ are shown in Fig. 4(a–c). Corresponding noisy images for $\sigma = 0.4$ are shown Fig. 4(d–f).



(a) True Label Map : $l = 1$ (b) True Label Map : $l = 3$ (c) True Label Map : $l = 5$

(d) Noisy Image: $l = 1, \sigma = 0.4$ (e) Noisy Image: $l = 3, \sigma = 0.4$ (f) Noisy Image: $l = 5, \sigma = 0.4$

FIG. 4. *Single realizations of ground truth label maps and corresponding noisy images generated from Matérn processes with length-scale parameter varied between 1, 3 and 5.*

For each setting of Matérn length scale l we generated 40 ground-truth binary label maps, and for each binary map we generated 5 noisy images at each noise level σ. Next, for every realized pair of binary and noisy images (z, y), the AMF approximating distribution Q was computed by solving the ROF equations on the logit maps of y. The original probability model P of Eqn. (5.1) was also explored using Gibbs sampling with 5 chains of $N = 10^5$ particles each, temperature $T = 1$ and thinning factor=0.1. The temperature parameter, which controls the scale of the sampling distribution, is needed because the probability distribution P is known only up to to a scale factor (i.e. the partition function). Therefore, the Gibbs probability of $z_i = 1$ is $\exp(-e_1/T)/(\exp(-e_1/T) + \exp(-e_0/T))$, where e_1 and e_0 are the energies corresponding to $z_i = 1$ and $z_i = 0$ respectively. Convergence was tested using the Gelman and Rubin diagnostic [25] resulting in approximately 2×10^4 particles being retained. Based on these Monte-Carlo particles, the following statistics were calculated for each realized image pair (z, y):

- *The correlation coefficient between the probability masses of each particle according to P and Q.* Note that although both P and Q are known only up to scales, it does not affect the correlation coefficient computation. As shown in Fig. 5 we see a strong correlation between the label map probabilities as estimated by AMF and the original model. This implies that the AMF model is a good approximation to the original probability model. However the correlation seems to reduce with increase in l and σ, implying that smoother images are harder to approximate - probably because of an increase of non-local interactions that cannot be well approximated by the mean-field distribution and that increasing noise causes greater mis-approximation.
- *The mean area of the label map estimated for P from the Gibbs samples and for Q by closed-form evaluation.* As shown in Fig. 6 the AMF model appears to underestimate the foreground's mean-area when it is less than 50% of the full image, but this underestimation improves as the foreground fraction

increases. Nevertheless there is good agreement, in terms of trend, between the mean area as estimated by the AMF model (in closed form) and the original probability model (via Gibbs sampling).

- *The variance in the area of the label map, again estimated for P from the Gibbs samples and for Q by closed-form evaluation.* As seen in Fig. 7, the second order statistics are not captured well by the AMF when compared to the second order statistics of the original model (as assessed by Gibbs sampling); especially for images with larger levels of smoothness. This is not surprising given that the mean-field approximation does not capture higher order interactions of the random field.

In summary, the posterior distribution of the AMF model correlates well with the posterior distribution obtained by Gibbs sampling. The obtained segmentation areas for the AMF model have the same trend as for Gibbs sampling, but tend to underestimate the segmentation area. As expected, higher-order statistics are not captured well due to the simplicity of the factorized variational distribution Q of the AMF model.

5.4. Assessment of AMF on Real Data. We illustrate the behavior of the AMF approach on real images qualitatively in Section 5.4.1 and quantitatively in Section 5.4.2. Our goal in this section is not to beat state-of-the art segmentation methods for our example segmentation applications (which may for example, use shape models or more sophisticated machine learning approaches to improve segmentation results), but to illustrate the AMF approach in the context of challenging datasets. Note, however, that the AMF model can be based on any foreground and background likelihood map. Therefore, it is able to augment other more sophisticated pre-processing to obtain foreground and background probabilities.

5.4.1. Qualitative Assessment of AMF on Real Data. We use ultrasound images of the prostate and the heart as well as an image of Fabio [40] to demonstrate the behavior of AMF under different levels of regularization. We limit ourselves to simple intensity distributions for the Fabio and the heart ultrasound image. We use a classifier supporting probabilistic outputs based on image intensities for the prostate example. Image size for the prostate example is 257 by 521 pixels, for the heart example 314 by 350 pixels, and for the Fabio image 253 by 254 pixels.

Fig. 8 shows an ultrasound image of the heart (left), an expert segmentation into blood pool, myocardium, and valves (middle) and the intensity distribution for the blood pool and outside the heart (right). These intensity distributions clearly overlap. We initialized the AMF model with this user-defined intensity distribution by sampling from the image followed by kernel-density estimation of the intensities. We re-estimated the intensity-distributions during the optimization. Specifically, given an intensity distribution, we compute the AMF solution, from that we obtain the binarized MAP solution that we use to re-estimate the intensity distributions using kernel-density estimation. We alternate AMF solution and density estimation to convergence. Fig. 9 shows the results of the AMF model for the estimation of label probabilities. The intensity ambiguity is captured in the estimated label probabilities of the AMF model. Regularization behaves as expected: low regularization results in noisy label probability maps. Moderate to high regularization allows capturing of the blood pool (for the MAP solution) while declaring other regions ambiguous or low-probability. Very large regularization declares the full image ambiguous, as expected, because the model will, in this case, prefer overly large segmentation regions.

Fig. 10 shows an ultrasound image of the prostate (left) and the corresponding

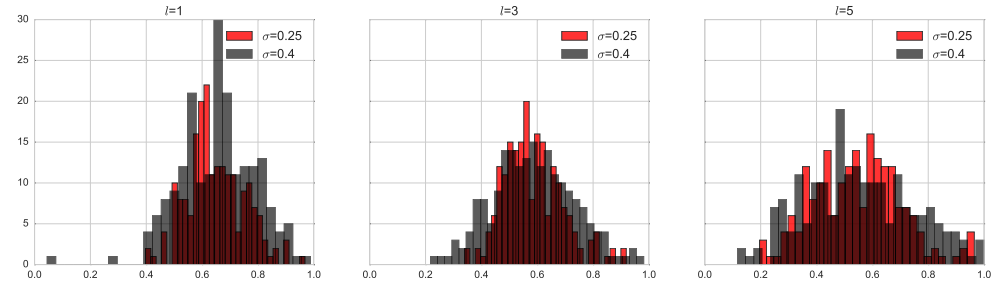


FIG. 5. *Histograms of the correlation coefficient between the posterior probability of the Gibbs samples as measured by P and Q. Each histogram is across the realizations of the synthetic binary maps and noisy images, i.e., one correlation coefficient per pair, for the various settings of Matérn length scale parameter l and image noise σ.*

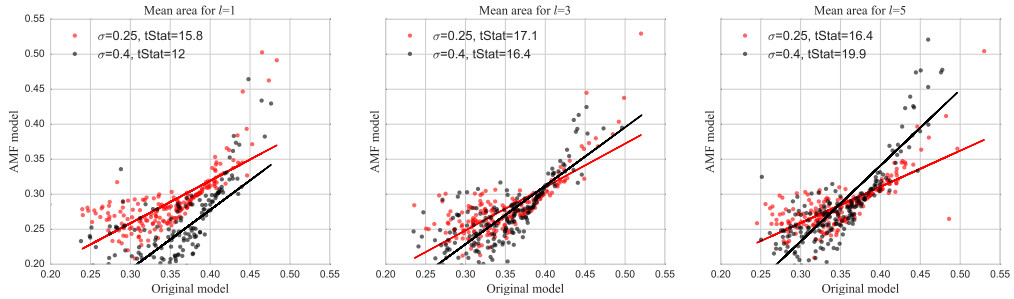


FIG. 6. *Scatter plots of the mean foreground area (as a fraction of total area) as measured under P (via Gibbs sampling) and Q (in closed form). Each point is one realization of the synthetic binary maps and noisy images for the various settings of Matérn length scale parameter l and image noise σ.*

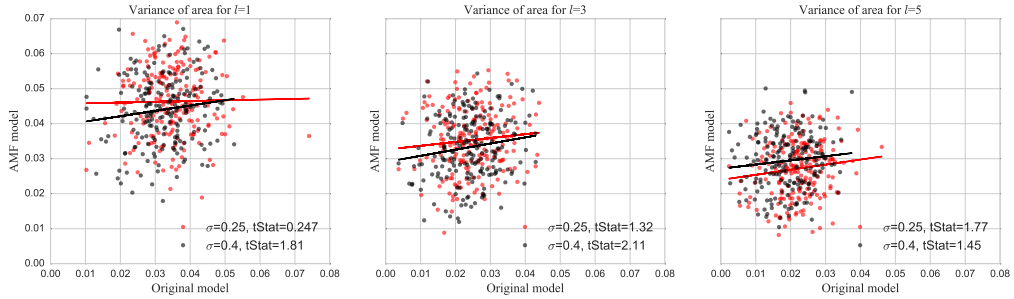


FIG. 7. *Scatter plots of the variance of the fractional foreground area as measured under P (via Gibbs sampling) and Q (in closed form). Each point is one realization of the synthetic binary maps and noisy images for the various settings of Matérn length scale parameter l and image noise σ.*

results of an experimental prostate segmentation system (right). The prototype system analyzed Radio Frequency (RF) ultrasound data using deep learning and random forest classification to generate label probabilities. Alternating optimization, as in the heart example, was not used. Fig. 11 shows the results of the AMF model. The same conclusions as for the heart example apply. More regularity yields cleaner looking probability images as the AMF smooths the probability field as expected because of the connection to the ROF model. Changes are not as drastic as for the heart example

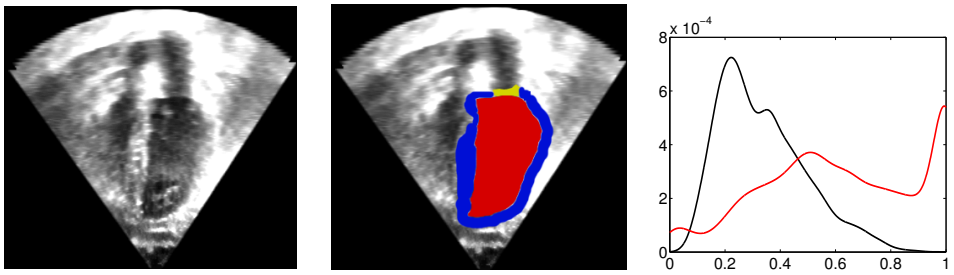


FIG. 8. *Left: ultrasound image of the heart. Middle: ultrasound image of the heart with overlaid expert segmentations of blood pool (red), myocardium (blue) and valves (yellow). Right: intensity distributions for the blood pool (red) and the areas outside of the heart (black) for the intensity-normalized image ($I \in [0, 1]$). Intensity distributions clearly overlap making an intensity-only segmentation challenging.*

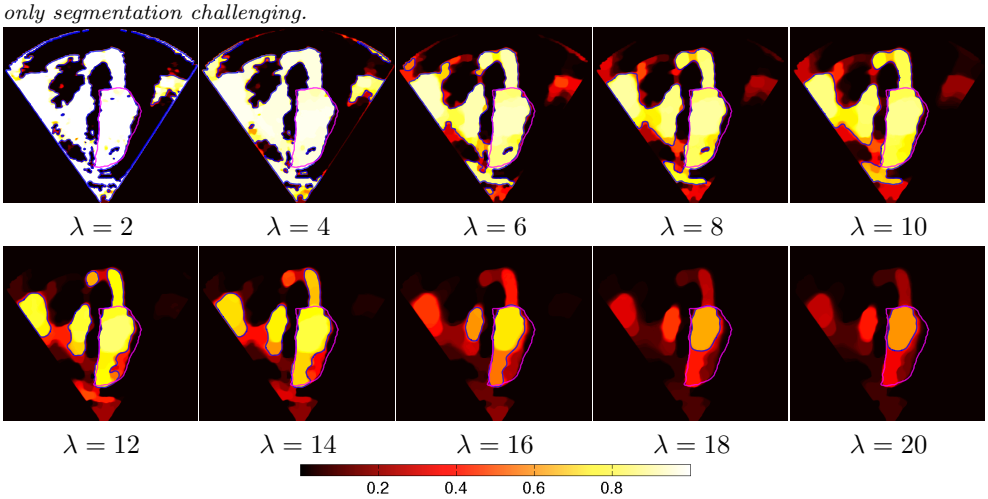


FIG. 9. *Intensity-based segmentation results of the heart from an ultrasound image for the AMF model. Increased regularization captures increasingly consistent regions. Moderate to high regularization retains high probabilities of the blood pool while estimating low probabilities for the surroundings. Very large regularization yields ambiguous label probabilities throughout the complete image. Magenta contour indicates expert segmentation of the blood-pool, blue contour indicates the* 0.5 *probability isocontour of the AMF solution.*

as the initial probability map is already substantially more regular.

Fig. 12 show the original Fabio image including its segmentations based on a modified version of Otsu thresholding (where foreground and background classes can have distinct means *and* standard deviations) and the corresponding intensity histogram. This image can be separated reasonably well using intensity information alone. Fig. 13 shows the corresponding AMF results. We obtained these results by initializing AMF using the modified Otsu-thresholding procedure and then followed the same alternating optimization approach as for the heart ultrasound segmentation. Clearly, larger values for the regularization parameter λ put the emphasis on larger image structures.

These experiments show that the AMF model (i) results in label probabilities which are spatially smooth (as expected due to the connection to the ROF model), (ii) exhibits a balancing effect between local label likelihood and spatial regularization, and (iii) tends to more uncertain label assignments for strong spatial regularization.

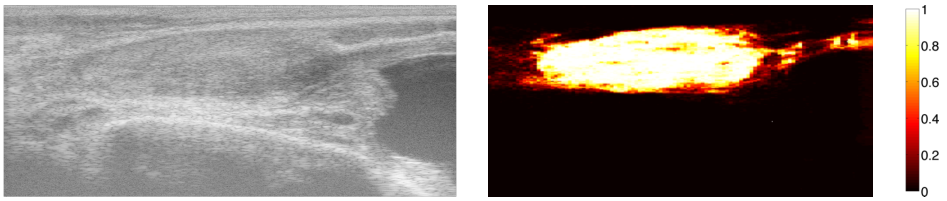


FIG. 10. *Left: ultrasound image of the prostate. Right: prostate probability map obtained by a machine-learning approach.*

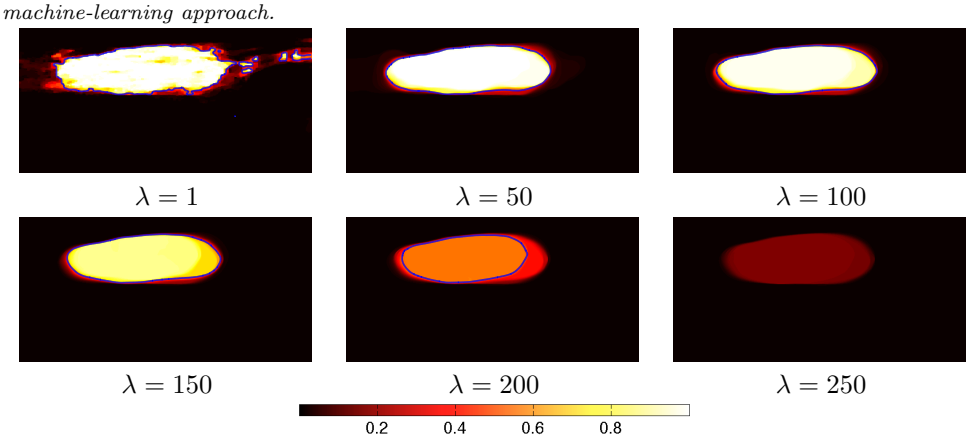


FIG. 11. *Probability-map-based segmentation results of the prostate from an ultrasound image for the AMF model. Input to the AMF is the prostate probability map of Fig. 10(right). Increased regularization captures increasingly consistent regions. Moderate to high regularization retain high probabilities of the prostate while estimating low probabilities for the surroundings. Very large regularizations yield ambiguous label probabilities throughout the complete image. Blue contour indicates the* 0.5 *probability isocontour of the AMF solution.*

5.4.2. Quantitative Assessment of AMF on Real Data. For quantitative analysis, we applied AMF to the segmentation benchmark data (`icgbench`) of Santner et al. [52]. This benchmark dataset consists of 158 natural color images (391 by 625 pixels). For each image a manual segmentation is available. Furthermore, each image contains seed regions for the objects to be segmented. In total there are 262 seed regions and 887 objects. As proposed by Santner et al. [52], we train a random forest (using Matlab's `TreeBagger` function) for each image given pixel color information in image areas defined by user-provided seed locations dilated by a disk structural element of radius of 9 pixels. Each random forest consists of 100 trees, λ was set to 10 for all the experiments; the models were trained on local CIElab color features. Once trained on the seeds, the resulting random forest classifier is applied to the full images generating noisy label probabilities. The mean computation time for an AMF segmentation was 22.1s for the RGB color images of the `icgbench` dataset using a Matlab CPU implementation on a 2GHz Intel Xeon, E5405. The computer had 8 cores, but the code was not explicitly multi-threaded (beyond what Matlab multi-threads automatically). As `icgbench` is a dataset for multi-label segmentation but our current AMF model only supports binary segmentation tasks[3], we investigate two different segmentation approaches:

[3]A multi-label extension is likely possible, but it remains to be investigated if connections to the ROF and the CV models can still be made.

FIG. 12. *Left: Fabio image. Middle: Otsu-thresholded Fabio image. Right: intensity distributions for the intensity-normalized image ($I \in [0,1]$) based on the classes determined by Otsu thresholding.*

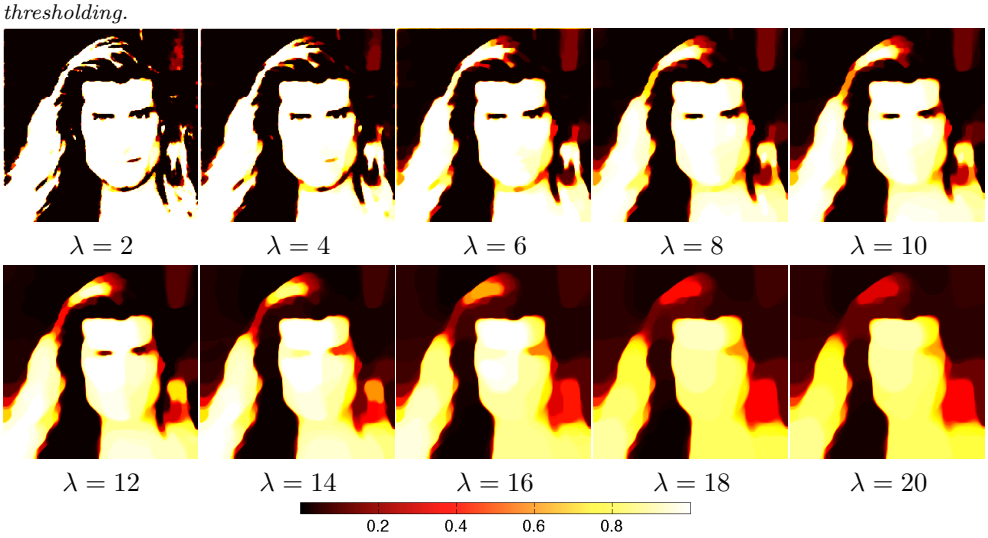


FIG. 13. *Intensity-based segmentation results for the Fabio image for the AMF model.*

- *Individual Binary Segmentations:* For a given image we create binary segmentations by considering one class as the foreground and all other classes as the background.
- *Quasi-Multi-Label Segmentation:* Individual binary segmentations do not guarantee that local label probabilities over all classes sum up to one. Hence, we project the local label probabilities obtained from the individual binary segmentations onto the probability simplex. We used the standard Euclidean projection [19, 38] onto the simplex though other approaches could be used as well [48, 1].

Fig. 14(left) compares the obtained Dice scores over all 887 individual object segmentations based on the random forest and based on AMF applied to the random forest label probabilities. The Dice score between two sets S_1 and S_2 is defined as

$$\text{Dice}(S_1, S_2) \triangleq \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}. \tag{5.2}$$

To evaluate image segmentations, S_1 and S_2 correspond to sets of object pixels which are the most likely for a given object class label (*i.e.*, foreground and background).

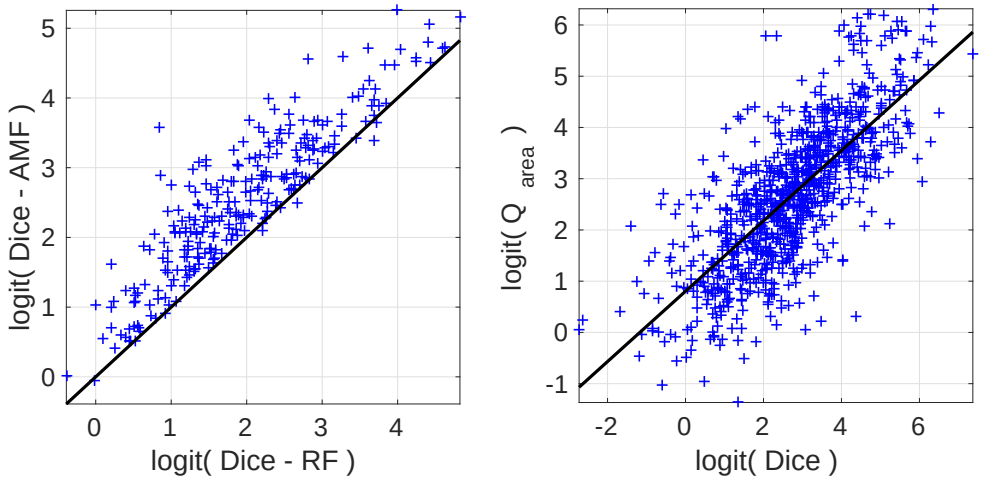


FIG. 14. *Left: Scatter plot for Dice segmentation scores for all the objects of the* `icgbench` *database. Comparison between obtained Dice scores through the random forest (RF) and after applying the AMF model (AMF). Values are logit transformed before plotting for better visualization:* $\mathit{logit}(p) = \ln(p/(1-p))$. *In the vast majority of the cases, AMF improves the Dice score. Line indicates equal values for RF and AMF model, i.e., values above the line indicate a better performance of the AMF model compared to the RF. Right: Scatter plot between Dice segmentation score and area-normalized posterior approximation Q of the AMF. Values are also logit transformed for better visualization. High Dice scores are generally related to high Q values. A clear linear trend is visible for the logit-transformed variables. Line indicates a least-squares fit to the logit transformed Q/Dice value pairs. Sample* $(p, \mathit{logit}(p))$ *pairs are as follows:* $(0.01, -4.60)$, $(0.25, -1.10)$, $(0.5, 0)$, $(0.75, 1.10)$, $(0.9, 2.20)$, $(0.99, 4.60)$.

AMF clearly improves the segmentations generated by the random forest. The mean Dice score (with standard deviation in parentheses) for the individual segmentations over all images is 0.82(0.18) for the random forest, which are significantly worse ($p < 10^{-10}$) according to a one-sided paired t-test) than individual binary AMF segmentations, whose mean Dice score is 0.88(0.15). The quasi-multi label AMF approach further improves the mean Dice score to 0.89(0.14). Computing multi-label Dice-scores[4] for all the images results in a mean Dice score of 0.84(0.11) for the random forest and 0.90(0.09) for the quasi-multi-label AMF segmentation, which is significantly better ($p < 1e - 10$ according to a one-sided paired t-test) and matches the Dice score obtained by Santner et al. [52] when using the same features.

Not only is our method simpler than the approach by Santner et al. [52], which uses a sophisticated random forest implementation coupled with a true multi-label segmentation approach (*i.e.*, all labels are jointly considered during the segmentation and not in a one-versus-all-other classes fashion as in our approach), but our method also complements the MAP solution with posterior label probabilities, which can be used to quantitatively assess the confidence in the segmentation. A possible confidence

[4]We compute the multi-label Dice score as the mean over the *individual* Dice scores for the individual binary segmentations for an image. Hence, we obtain *one* multi-label Dice score per *image*, but as many individual Dice scores as there are objects in an image.

measure is to compute an *area-normalized* approximation of the posterior

$$Q(\mathrm{z};\ \theta) \triangleq \prod_{i \in \mathbb{I}_{\mathcal{X}}} \theta_i^{\mathrm{z}_i} \left(1-\theta_i\right)^{1-\mathrm{z}_i} \tag{5.3}$$

$$= \exp\left\{ \sum_{i \in \{i:\mathrm{z}_i=1\}} \ln(\theta_i) + \sum_{i \in \{i:\mathrm{z}_i=0\}} \ln(1-\theta_i) \right\}. \tag{5.4}$$

Area-normalization is useful as object sizes in the `icgbench` dataset vary greatly. Specifically, we define the area-normalized form of $Q(\cdot;\cdot)$ as

$$Q_{\text{area}}(\mathrm{z};\ \theta) \triangleq \exp\left\{ \frac{1}{|\{i : \mathrm{z}_i = 1\}|} \sum_{i \in \{i:\mathrm{z}_i=1\}} \ln(\theta_i) + \frac{1}{|\{i : \mathrm{z}_i = 0\}|} \sum_{i \in \{i:\mathrm{z}_i=0\}} \ln(1-\theta_i) \right\}. \tag{5.5}$$

We use the MAP solution of the AMF, z_{map}, to evaluate Q_{area}. For a binary θ, *i.e.*, no uncertainty in the inferred binary segmentation, $Q_{\text{area}}(\mathrm{z}_{map};\ \theta) = 1$. The value of $Q_{\text{area}}(\mathrm{z}_{map};\ \theta)$ decreases if θ is not binary indicating uncertainty in the inferred segmentation. In comparison to the approximate posterior Q, this area-normalized measure allows us to assess uncertainty of objects independent of their size. For $Q_{\text{area}}(\mathrm{z}_{map};\ \theta)$ to be a useful measure of segmentation quality, it should be high for high Dice scores and conversely low for low Dice scores. The scatter plot between Dice scores and $Q_{\text{area}}(\mathrm{z}_{map};\ \theta)$ in Fig. 14(right) shows that this is indeed frequently the case. Hence $Q_{\text{area}}(\mathrm{z}_{map};\ \theta)$ can serve as a measure of segmentation confidence in the absence of manual segmentations.

To gain a deeper understanding of the Q_{area} measure, it is instructive to review cases where Q_{area} seems unrelated to the Dice score. Fig. 15 shows a case with very high Q_{area}, but low Dice score, caused by a very confident, but incorrect output of the random forest from which the AMF cannot recover. Fig. 16 shows a case with very low Q_{area} but high Dice score. Here, the segmentation is good, but our approach is not confident as other regions have similar color. Fig. 17 and Fig. 18 show examples where segmentations receive both high Dice and Q_{area} scores, indicating high quality segmentations which also have high segmentation confidence according to Q_{area}.

In summary, the AMF model shows good segmentation performance across a large set of natural images. Furthermore, the posterior distribution on labels carries useful information as it can provide a proxy for likely segmentation quality.

6. Conclusions. We described a method for binary image segmentation which allows efficient estimation of approximate label probabilities through a VMF approximation. We carefully analyzed the theoretical properties of the model and tested its behavior on synthetic and real datasets. A particularly useful feature of our model is that it has strong connections to the Chan-Vese segmentation model and the ROF image-denoising model. Our method can therefore be implemented using off-the-shelf solvers of the ROF model. This simple and efficient way to compute solutions makes AMF an attractive alternative to Chan-Vese-like approaches, which, unlike AMF, do not compute posterior distributions on labels. A current drawback of our method is its binary formulation. Nevertheless, our approach can be used for multi-label segmentation by converting multi-label problems to multiple binary segmentations. A truly multi-label formulation of AMF is outside the scope of this paper, but should be investigated in future work. It will be interesting to see if connections to the Chan-Vese and the ROF model can also be established in a multi-label VMF approach.

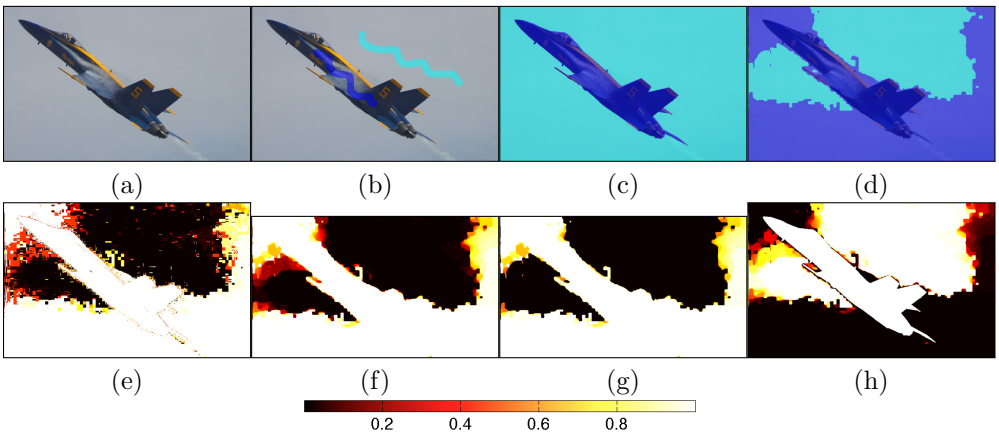


FIG. 15. *Unusual case: Segmentation with high confidence, but low Dice score indicating a segmentation of low quality. (a) Original Image; (b) Seeds to train the random forest; (c) Expert segmentation; (d) AMF segmentation; (e) label probabilities for plane object computed by random forest ; (f) AMF-computed label probabilities for the plane object; (g) masked AMF-computed label probabilities, only showing areas where the plane object is most probable ; (h) AMF-computed label probabilities for the correct expert labels at each location (white image, $\theta = 1$ would be a perfect result). As the color values for the plane seeds (dark blue) are similar to regions in the sky, the random forest classifier (e) is overly confident from which the AMF (f) cannot recover. Hence, there is poor overlap between the resulting segmentation (d and h) and the expert labeling (c). At the same time, the overall confidence for this example is high due to the high certainty of the random forest approach. The Dice score for the quasi-multi-label AMF and for the binary AMF is 0.49. The mean Q_{area} score for both approaches is* 0.94.

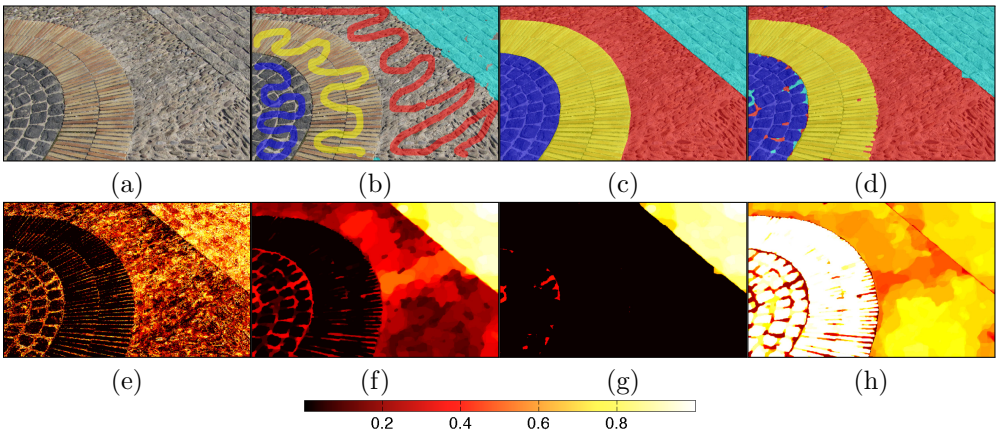


FIG. 16. *Unusual case: Segmentation with low confidence, but high Dice score indicating a segmentation of high quality. (a) Original Image; (b) Seeds to train the random forest; (c) Expert segmentation; (d) AMF segmentation; (e) random forest label probabilities for cobble-stone object on the top-right; (f) corresponding label probabilities computed by AMF; (g) masked AMF-computed label probabilities, only showing areas where the cobble-stone object is the most probable; (h) AMF-computed label probabilities for the correct expert labels at each location (white image, $\theta = 1$ would be a perfect result). In this example, the seed points for the cobble-stone object (b; cyan) essentially fully segment the object of interest. However as the color values are ambiguous with respect to the other classes (in particular the red seed label) the overall segmentation result is not highly confident (f and g) resulting in a lower Q_{area} score. However, the most probable labelings also agree with the experts' opinion (d and h). The Dice score for the quasi-multi-label AMF is 0.97 and for the binary AMF 0.93. The mean Q_{area} scores are 0.71 and 0.67 respectively.*

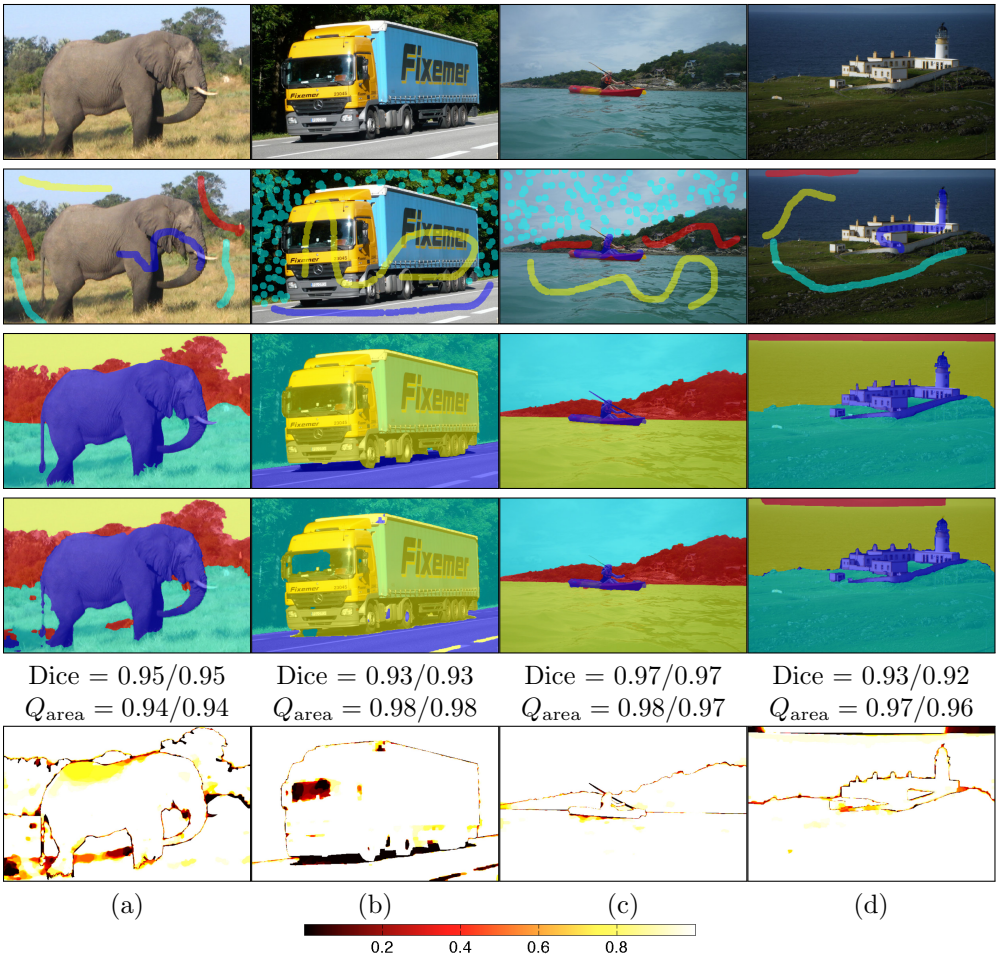


FIG. 17. *Sample segmentation results for highly confident high quality segmentations. Top row: original images; 2nd row: seeds used for training the random forest; 3rd row: expert segmentations; 4th row: AMF segmentation result; last row: label probabilities computed by AMF with respect to the object segmented by the expert (i.e., given an expert label the corresponding probability for that label as computed by the AMF is displayed; a perfect result would be a totally white image). Dice scores for the quasi-multi label approach applied to the AMF segmentation, majority voting (first value), and the mean for all binary segmentations for a given image respectively. Q_{area} scores are the means over all the segmented objects for the quasi-multi-label approach (first value) and the binary segmentation approach (second value).*

7. Acknowledgements. This work was supported by NIH grants P41RR013218, P41EB015898, P41EB015902, NIAID R24 AI067039, R01CA111288, R01 HL127661, K05 AA017168 and NSF grants EECS-1148870, EECS-0925875. We also gratefully acknowledge Tassilo Klien and Andrey Fedorov for providing prostate data and probability maps, and Dr. Theodore Abraham for the heart data.

Appendix A.

We compare the approximated and exact distributions, $Q(\mathrm{z};\theta)$ and $P(\mathrm{z}|y)$, respectively, for general realizations of z. Because the normalizer for $P(\mathrm{z}|y)$ is not available, and for convenience, we will compare $\ln \frac{P(\mathrm{z}|y)}{P(\mathrm{z}_0|y)}$ and $\ln \frac{Q(\mathrm{z};\theta)}{Q(\mathrm{z}_0;\theta)}$, where z_0 is

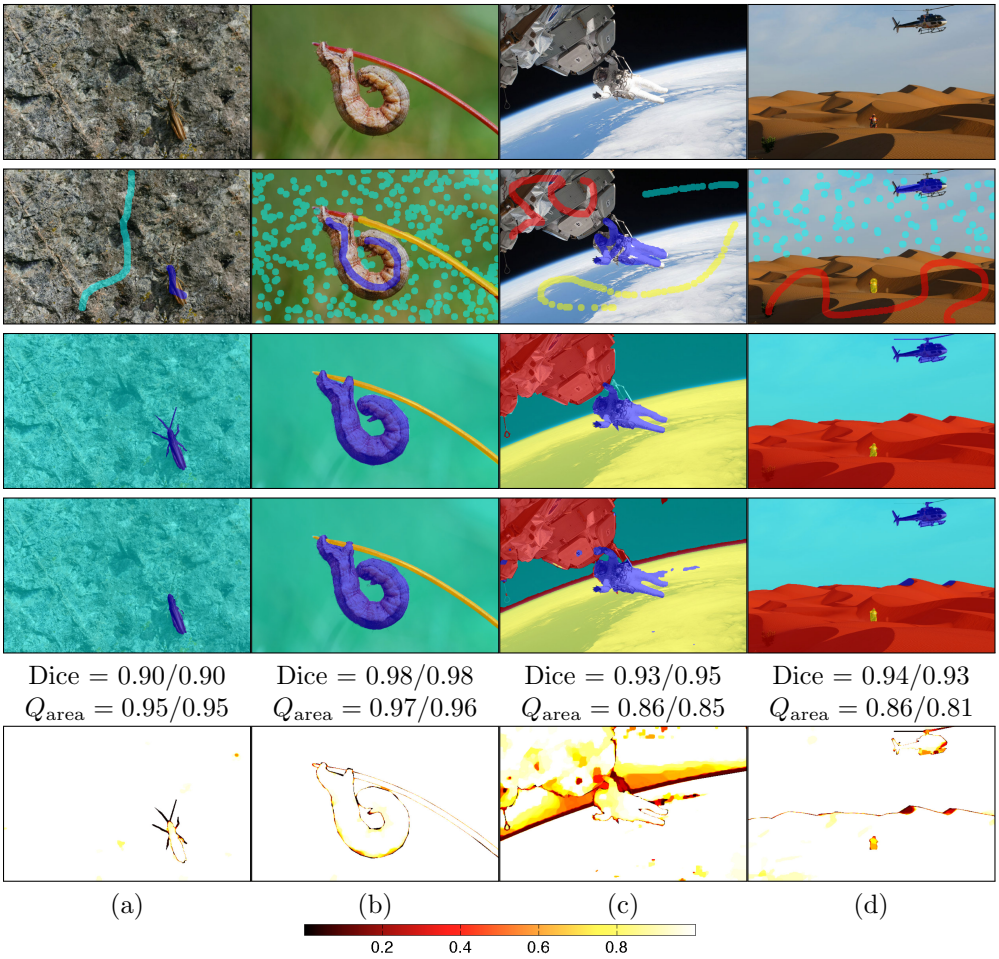


FIG. 18. *Sample segmentation results for highly confident high quality segmentations. Top row: original images; 2nd row: seeds used for training the random forest only; 3rd row: expert segmentation; 4th row: AMF segmentation result; last row: label probabilities of AMF with respect to the object segmented by the expert (i.e., given an expert label the corresponding probability for that label as computed by the AMF is displayed; a perfect result would be a totally white image). Dice scores for the quasi-multi label approach applied to AMF segmentation, majority voting (first value) and the mean for all binary segmentations for a given image respectively. Q_{area} scores are the means over all the segmented objects for the quasi-multi-label approach (first value) and the binary segmentation approach (second value).*

the most probable realization under Q.

For calculating the log probability ratio of P, we return to the original probability model. From Eqn. (2.3) - Eqn. (2.5),

$$\ln P(\mathrm{z}|y) = \sum_{i \in \mathbb{I}_{\mathcal{X}}} \mathrm{z}_i \psi_i - \lambda L(\mathrm{z}) + \text{const} \tag{A.1}$$

$$\approx v^{-1} \int_{\mathcal{X}} \mathrm{z}(x)\psi(x)dx - \lambda L(\mathrm{z}) + \text{const}\,. \tag{A.2}$$

Then the log probability ratio for the exact posterior, P, is

$$\ln \frac{P(\mathrm{z}|y)}{P(\mathrm{z}_0|y)} \approx v^{-1} \int_{\mathcal{X}} (\mathrm{z}(x) - \mathrm{z}_0(x))\psi(x)dx - \lambda L(\mathrm{z}) + \lambda L(\mathrm{z}_0) \ .$$

Working towards the probability ratio for the AMF approximate posterior, Q, using Eqn. (2.7) and using a similar technique,

$$\ln Q(\mathrm{z};\theta) \approx v^{-1} \int_{\mathcal{X}} \mathrm{z}(x)\phi(x)dx + \text{const} \ .$$

Here it is easy to see that the most probable realization under Q is bounded by the zero level-set of ϕ.

We may now write the log probability ratio for Q,

$$\ln \frac{Q(\mathrm{z};\theta)}{Q(\mathrm{z}_0;\theta)} \approx v^{-1} \int_{\mathcal{X}} (\mathrm{z}(x) - \mathrm{z}_0(x))\phi(x)dx \ .$$

Subtracting the two probability ratios,

$$\ln \frac{P(\mathrm{z}|y)}{P(\mathrm{z}_0|y)} - \ln \frac{Q(\mathrm{z};\theta)}{Q(\mathrm{z}_0;\theta)} = v^{-1} \int_{\mathcal{X}} (\mathrm{z}(x) - \mathrm{z}_0(x))(\psi(x) - \phi(x))dx - \lambda L(\mathrm{z}) + \lambda L(\mathrm{z}_0) \ . \tag{A.3}$$

We now make use of the AMF equation, $\phi(x) - \psi(x) - v\lambda\kappa(\phi(x)) = 0$, to establish relationships among the log probability ratios of p and q. We obtain

$$\ln \frac{P(\mathrm{z}|y)}{P(\mathrm{z}_0|y)} - \ln \frac{Q(\mathrm{z};\theta)}{Q(\mathrm{z}_0;\theta)} \tag{A.4}$$

$$= \lambda \left[- \int_{\mathcal{X}} (\mathrm{z}(x) - \mathrm{z}_0(x))\kappa(\phi(x))dx - L(\mathrm{z}) + L(\mathrm{z}_0) \right] \tag{A.5}$$

$$= \lambda \Big[- \int_{\mathcal{X}:\mathrm{z}(x)=1} \nabla \cdot \left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|} \right) dx \tag{A.6}$$

$$+ \int_{\mathcal{X}:\mathrm{z}_0(x)=1} \nabla \cdot \left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|} \right) dx - L(\mathrm{z}) + L(\mathrm{z}_0) \Big] \tag{A.7}$$

$$= \lambda \Big[- \int_{c(s)} N(x) \cdot \left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|} \right) ds \tag{A.8}$$

$$+ \int_{c_0(s)} N(x) \cdot \left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|} \right) ds - L(\mathrm{z}) + L(\mathrm{z}_0) \Big]. \tag{A.9}$$

The last two lines use the divergence theorem; $c(s)$ is the boundary of $\mathrm{z}(x)$ oriented so that the outward normal points from $\mathrm{z}(x) = 1$ towards $\mathrm{z}(x) = 0$, and similarly for $c_0(x)$ and $\mathrm{z}_0(x)$. $N(x)$ is the outward normal vector to the curve in question (see Fig. 19).

Then

$$\ln \frac{P(\mathrm{z}|y)}{P(\mathrm{z}_0|y)} - \ln \frac{Q(\mathrm{z};\theta)}{Q(\mathrm{z}_0;\theta)} = \lambda \left[\int_{c(s)} \beta(x)ds - \int_{c_0(s)} \beta(x)ds - L(\mathrm{z}) + L(\mathrm{z}_0) \right] \ , \tag{A.10}$$

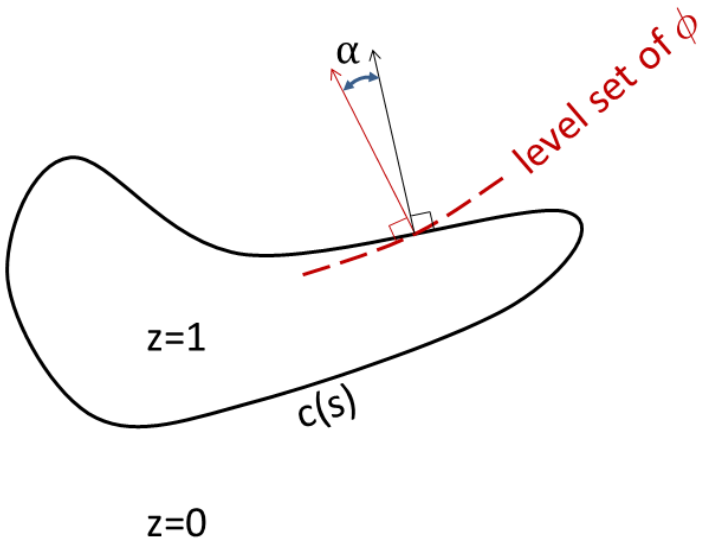


FIG. 19. *Level sets and normals.*

where

$$\beta(x) \doteq N(x) \cdot \left(\frac{-\nabla\phi(x)}{|\nabla\phi(x)|} \right)$$

is the dot product of two unit vectors, the outward normal to the curve and the negative of the direction of the gradient of ϕ.

On the curve c_0, $\beta(x) = 1$, because the boundary of z_0 is a level-set of $\phi(x)$. In that case the second and fourth terms cancel. Re-writing the third term as an integral over c,

$$\text{(A.11)} \quad \ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z;\theta)}{Q(z_0;\theta)} = \lambda \left[\int_{c(s)} \beta(x) ds - \int_{c(s)} 1 ds \right] = \lambda \int_{c(s)} (\beta(x) - 1) ds \ .$$

Because $\beta(x)$ is the dot product of two unit vectors, we may write $\beta(x) = \cos(\alpha(x))$, where α is the angle between the normal to the curve and the negative of the gradient direction of $\phi(x)$ (see Fig. 19). Then, using $\cos(\alpha) - 1 = -2\sin^2(\frac{\alpha}{2})$,

$$\ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z;\theta)}{Q(z_0;\theta)} = -2\lambda \int_{c(s)} \sin^2 \left(\frac{\alpha(x)}{2} \right) ds \ .$$

Summarizing the comparison of the probability ratios of the exact and approximate distributions, P and Q, respectively we see the following:

- For realizations that are bounded by level-sets of ϕ, α is zero, so the probability ratios agree.
- For realizations whose boundaries are in direction "close" to level-sets of ϕ, the probability ratios approximately agree (the disagreement is quadratic in α).
- For curves where α is not small, the probability ratio for Q will be larger than for P, *i.e.*, Q underestimates the length penalty of P.

We saw above that the zero level-set of ϕ is the boundary of the most probable realization under the approximate distribution, $Q(\mathrm{z};\theta)$ (and it is unique). Since the probability ratios agree for z_0 (a level set of ϕ), and the Q ratio upper-bounds the P ratio, we conclude that it is also the boundary of the MAP realization under $P(\mathrm{z}|y)$. In summary, z_0, whose boundary is the zero level-set of ϕ, satisfies

$$\mathrm{z}_0 \doteq H(\phi(x)) = \arg\max_z Q(\mathrm{z};\theta) = \arg\max_z P(\mathrm{z}|y) \ .$$

REFERENCES

[1] S. Andrews, N. Changizi, and G. Hamarneh, *The isometric log-ratio transform for probabilistic multi-label anatomical shape representation*, Medical Imaging, IEEE Transactions on, 33 (2014), pp. 1890–1899.

[2] B. Appleton and H. Talbot, *Globally minimal surfaces by continuous maximal flows*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28 (2006), pp. 106–118.

[3] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, Image Processing, IEEE Transactions on, 18 (2009), pp. 2419–2434.

[4] J. Besag, *On the statistical analysis of dirty pictures*, J Royal Stat Soc., Series B (Methodological), 48 (1986), pp. 259–302, http://www.jstor.org/stable/2345426.

[5] F. Bethuel and J.-M. Ghidaglia, *Geometry in Partial Differential Equations*, World Scientific, 1994, ch. 1, pp. 1–17.

[6] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[7] Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23 (2001), pp. 1222–1239.

[8] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher, *Fast global minimization of the active contour/snake model*, Journal of Mathematical Imaging and Vision, 28 (2007), pp. 151–167.

[9] M. Burger and S. Osher, *A guide to the TV zoo*, in Level Set and PDE Based Reconstruction Methods in Imaging, Springer, 2013, pp. 1–70.

[10] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic active contours*, International journal of computer vision, 22 (1997), pp. 61–79.

[11] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical imaging and vision, 20 (2004), pp. 89–97.

[12] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numerica, 25 (2016), pp. 161–319.

[13] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837.

[14] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM Journal on Scientific Computing, 20 (1999), pp. 1964–1977.

[15] T. F. Chan, B. Y. Sandberg, and L. A. Vese, *Active contours without edges for vector-valued images*, Journal of Visual Communication and Image Representation, 11 (2000), pp. 130–141.

[16] T. F. Chan and L. A. Vese, *Active contours without edges*, Image Processing, IEEE Transactions on, 10 (2001), pp. 266–277.

[17] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, 1987.

[18] J. Chang and J. W. F. III, *Efficient MCMC sampling with implicit shape representations*, in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 2081–2088.

[19] Y. Chen and X. Ye, *Projection onto a simplex*. arXiv:1101.6081v2, 2011.

[20] D. Cremers, M. Rousson, and R. Deriche, *A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape*, International journal of computer vision, 72 (2007), pp. 195–215.

[21] N. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, 1991.

[22] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the royal statistical society. Series B (methodological), (1977), pp. 1–38.

[23] S. S. Dragomir, *A converse of the Jensen inequality for convex mappings of several variables and applications*, Acta Mathematica Vietnamica, 29 (2004), pp. 77–88, http://arxiv.org/

abs/math/0304345.

[24] A. C. Fan, J. W. F. III, W. M. W. III, J. J. Levitt, and A. S. Willsky, *MCMC curve sampling for image segmentation*, in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007, Springer, 2007, pp. 477–485.

[25] A. Gelman and D. B. Rubin, *Inference from iterative simulation using multiple sequences*, Statistical Science, 7 (1992), pp. 457–472.

[26] T. Georgiou, O. Michailovich, Y. Rathi, J. Malcolm, and A. Tannenbaum, *Distribution metrics and image segmentation*, Linear algebra and its applications, 425 (2007), pp. 663–672.

[27] T. Goldstein and S. Osher, *The split Bregman method for l1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343.

[28] K. Held, E. R. Kops, B. Krause, W. W. III, R. Kikinis, and H. Müller-Gärtner, *Markov random field segmentation of brain MR images*, Medical Imaging, IEEE Transactions on, 16 (1997), pp. 878–886.

[29] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift für Physik A Hadrons and Nuclei, 31 (1925), pp. 253–258.

[30] M. G. Jameson, S. Kumar, S. K. Vinod, P. E. Metcalfe, and L. C. Holloway, *Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy*, Radiotherapy and Oncology, 112 (2014), pp. 332–336.

[31] T. Kapur, W. E. L. Grimson, R. Kikinis, and W. M. W. III, *Enhanced spatial priors for segmentation of magnetic resonance imagery*, in Medical Image Computing and Computer-Assisted Intervention MICCAI'98, Springer, 1998, pp. 457–468.

[32] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, International journal of computer vision, 1 (1988), pp. 321–331.

[33] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, *Gradient flows and geometric active contour models*, in Computer Vision, 1995. Proceedings., Fifth International Conference on, IEEE, 1995, pp. 810–815.

[34] V. Kolmogorov and R. Zabih, *What energy functions can be minimized via graph cuts?*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26 (2004), pp. 147–159.

[35] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer, 2009.

[36] R. Malladi, J. A. Sethian, and B. C. Vemuri, *Shape modeling with front propagation: A level set approach*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 17 (1995), pp. 158–175.

[37] S. Martin, C. Johnson, M. Brophy, D. A. Palma, J. L. Barron, S. S. Beauchemin, A. V. Louie, E. Yu, B. Yaremko, B. Ahmad, and G. B. Rodrigues, *Impact of target volume segmentation accuracy and variability on treatment planning for 4D-CT-based non-small cell lung cancer radiotherapy*, Acta Oncologica, 54 (2015), pp. 322–332.

[38] C. Michelot, *A finite algorithm for the projection of a point onto the canonical simplex of* $\mathbb{R}^n$, Journal of Optimization Theory and Applications, 50 (1986), pp. 195–200.

[39] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on pure and applied mathematics, 42 (1989), pp. 577–685.

[40] D. Needell and R. Ward, *Stable image reconstruction using total variation minimization*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1035–1058.

[41] M. Niethammer and C. Zach, *Segmentation with area constraints*, Medical image analysis, 17 (2013), pp. 101–112.

[42] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer, 2003.

[43] N. Otsu, *A threshold selection method from gray-level histograms*, Automatica, 11 (1975), pp. 23–27.

[44] N. Paragios and R. Deriche, *Geodesic active regions: A new framework to deal with frame partition problems in computer vision*, Journal of Visual Communication and Image Representation, 13 (2002), pp. 249–268.

[45] K. Petersen, M. Nielsen, and S. S. Brandt, *A static SMC sampler on shapes for the automated segmentation of aortic calcifications*, in Computer Vision–ECCV 2010, Springer, 2010, pp. 666–679.

[46] T. Plefka, *Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model*, Journal of Physics A: Mathematical and general, 15 (1982), p. 1971.

[47] T. Pock, M. Unger, D. Cremers, and H. Bischof, *Fast and exact solution of total variation models on the GPU*, in Computer Vision and Pattern Recognition Workshops, CVPRW'08, IEEE, 2008, pp. 1–8.

[48] K. M. Pohl, J. Fisher, S. Bouix, M. Shenton, R. W. McCarley, W. E. L. Grimson, R. Kikinis, and W. M. W. III, *Using the logarithm of odds to define a vector space on*

probabilistic atlases, Medical Image Analysis, 11 (2007), pp. 465–477.

[49] K. M. Pohl, J. Fisher, W. E. L. Grimson, R. Kikinis, and W. M. W. III, *A Bayesian model for joint segmentation and registration*, NeuroImage, 31 (2006), pp. 228–239.

[50] K. M. Pohl, R. Kikinis, and W. M. W. III, *Active mean fields: Solving the mean field approximation in the level set framework*, in Information Processing in Medical Imaging, N. Karssemeijer and B. Lelieveldt, eds., vol. 4584 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2007, pp. 26–37, https://doi.org/10.1007/978-3-540-73273-0_3, http://dx.doi.org/10.1007/978-3-540-73273-0_3.

[51] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.

[52] J. Santner, T. Pock, and H. Bischof, *Interactive multi-label segmentation*, in Proceedings of the 10th Asian Conference on Computer Vision (ACCV), Queenstown, New Zealand, November 2010.

[53] G. Sapiro, *Geometric Partial Differential Equations and Image Analysis*, Cambridge, 2001.

[54] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, *A comparative study of energy minimization methods for Markov random fields with smoothness-based priors*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30 (2008), pp. 1068–1080, https://doi.org/10.1109/TPAMI.2007.70844.

[55] J. L. Troutman, *Variational calculus and optimal control: optimization with elementary convexity*, Springer Science & Business Media, 2012.

[56] M. W. Vannier, R. L. Butterfield, D. Jordan, W. A. Murphy, R. G. Levitt, and M. Gado, *Multispectral analysis of magnetic resonance images*, Radiology, 154 (1985), pp. 221–224.

[57] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM Journal on Scientific Computing, 17 (1996), pp. 227–238.

[58] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families, and variational inference*, Foundations and Trends in Machine Learning, 1 (2008), pp. 1–305.

[59] Y. Zhang, M. Brady, and S. Smith, *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm*, Medical Imaging, IEEE Transactions on, 20 (2001), pp. 45–57.