Medical Image Analysis 55 (2019) 181–196

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Estimating uncertainty in MRF-based image segmentation: A perfect-MCMC approach

Suyash P. Awate*, Saurabh Garg, Rohit Jena

Computer Science and Engineering Department, Indian Institute of Technology (IIT) Bombay, Mumbai, India

ARTICLE INFO

Article history:
Received 16 September 2018
Revised 19 April 2019
Accepted 30 April 2019
Available online 8 May 2019

Keywords:
Segmentation
uncertainty
Hidden MRF
Bayesian inference
EM
MCMC
Perfect/ exact sampling
brain
MRI
Multiatlas
Hippocampus
Tumor
Tissue
Lesion
Lobes

ABSTRACT

Typical methods for image segmentation, or labeling, formulate and solve an optimization problem to produce a single optimal solution. For applications in clinical decision support relying on automated medical image segmentation, it is also desirable for methods to inform about (i) the *uncertainty* in label assignments or object boundaries or (ii) alternate close-to-optimal solutions. However, typical methods fail to do so. To estimate uncertainty, while some Bayesian methods rely on simplified prior models and approximate variational inference schemes, others rely on sampling segmentations from the associated posterior model using (i) traditional Markov chain Monte Carlo (MCMC) methods based on Gibbs sampling or (ii) approximate perturbation models. However, in such typical approaches, in practice, the resulting inference or generated sample set are approximations that deviate significantly from those indicated by the true posterior. To estimate uncertainty, we propose the modern paradigm of *perfect MCMC sampling* to sample multi-label segmentations from *generic Bayesian Markov random field* (MRF) models, in finite time for exact inference. Furthermore, for exact sampling in generic Bayesian MRFs, we extend the theory underlying Fill's algorithm to *generic MRF models* by proposing a novel *bounding-chain* algorithm. On several classic problems in medical image analysis, and several modeling and inference schemes, results on simulated data and clinical brain magnetic resonance images show that our uncertainty estimates gain accuracy over several state-of-the-art inference methods.

1. Introduction

In applications related to clinical decision support relying on medical image analysis, exposing the *uncertainty* in automated image analyses (Folgoc et al., 2017) can lead to better informed decisions and better outcomes. For instance, estimating the uncertainty in automated segmentation results can be crucial for risk assessment and planning in clinical procedures, e.g., radiotherapy (Le et al., 2016) and neurosurgery (Alberts et al., 2016). It can also lead to increased reliability in clinical diagnoses and scientific studies. Popular segmentation methods, such as those relying on hidden Markov random field (MRF) modeling and subsequent optimization using graph cuts (Han et al., 2011) or expectation maximization (EM) (Zhang et al., 2001), typically produce a single optimal solution, failing to inform about uncertainty in (i) label assignments or object boundaries, and (ii) alternate close-to-optimal solutions.

For a very small class of MRF models that lend themselves to segmentation inference using graph-cut based optimization, efficient methods exist to exactly estimate a notion of label uncertainty (Kohli and Torr, 2008). However, their notion of label uncertainty cannot generalize to generic MRF models. In the context of image segmentation, a general notion of uncertainty can be considered to be the variance or unalikeability (Perry and Kader, 2005; Kader and Perry, 2007) in the label assignments stemming from the posterior distribution of the label image. For general MRFs, typical methods to estimate uncertainty either (i) infer approximate models to the posterior, from which sampling is easy or the variances can be estimated analytically, or (ii) use approximate sampling from the posterior. Examples of approximate models include those inferred using variational Bayesian (VB) methods, e.g., mean-field approximations, and using Gaussian-process (GP) models enforcing a simplified prior model. After fitting such simplified / approximate models to the true posterior, variances can typically be estimated analytically. Such approaches have been used in the context of voxel-labeling applications (Le et al., 2016). Examples of approximate sampling methods include traditional Markov chain Monte Carlo (MCMC) methods like Gibbs sampling

* Corresponding author.
E-mail address: suyash@cse.iitb.ac.in (S.P. Awate).

https://doi.org/10.1016/j.media.2019.04.014
1361-8415/

(Geman and Geman, 1984). Traditional MCMC approaches have been used in the context of object delineation applications (Fan et al., 2007). Some recent theoretical advances propose the perturb-and-MAP framework and a Gumbel perturbation model (GPM) (Papandreou and Yuille, 2011; Hazan et al., 2013) to exactly sample from MRF distributions. However, GPMs are practically intractable for large-sized MRFs, as observed in typical image segmentation applications and, thus, are approximated in practice (Alberts et al., 2016) that, we find, leads to loss of accuracy.

For uncertainty estimation in image registration, early methods (Kybic, 2010) use bootstrap data resampling to approximate the data distribution, instead of the posterior distribution of the labels. Recent pioneering methods (Folgoc et al., 2017) in this domain rely on MCMC sampling. However, these methods focus on the application of image registration (inferring continuous-valued deformation maps), while this manuscript focuses on a different problem, i.e., image segmentation (inferring discrete label images). Moreover, the framework proposed in this manuscript varies significantly from that proposed in the works related to image registration.

We introduce a novel paradigm for uncertainty estimation in image segmentation by relying on *perfect MCMC sampling*, in finite time, from *generic Bayesian MRF* models. We propose to sample label images from their true posterior distribution in two ways: (i) by combining coupling-from-the-past (CFTP) (Propp and Wilson, 1996) with the bounding-chain (BC) (Huber, 2004) scheme, called CFTP-BC, and, more importantly, (ii) by using a novel theoretical extension of Fill's algorithm (FA) (Fill, 1998) using the BC scheme, called FA-BC. We show advantages of FA-BC over CFTP-BC. We show that our perfect-MCMC sampling approach to estimate uncertainty actually performs superior to Bayesian modeling schemes that either simplify the prior model or perform approximate variational inference, even though the simplified / approximate models lead to analytical estimates of the uncertainty. Results on clinical brain magnetic resonance images from four classic applications, for segmenting subcortical structures, tumor, tissues, and lobes, show that our uncertainty estimates gain accuracy over several state-of-the-art methods.

This paper is organized as follows. Section 2 describes the related literature. Section 3 describes our proposed theoretical framework for perfect MCMC sampling to estimate uncertainty. Section 4.1 describes our proposed method for uncertainty estimation and the details of the probabilistic graphical models and inference strategies used by our framework, and other frameworks, for specific applications. Section 5 shows empirical results, quantitative and qualitative, on simulated and clinical data. Section 6 concludes the paper.

2. Related work

Probabilistic graphical models (Koller and Friedman, 2009) such as Ising models and Potts model are very important for image analysis and computer vision problems involving discrete labels such as image segmentation (Besag, 1974; Li, 2009). In general MRFs models on high-dimensional discrete spaces, exact probabilistic inference and maximum-likelihood estimation for parameters is often intractable, in which case random sampling methods are employed for inference.

Metropolis–Hastings (M–H) sampling (Hastings, 1970) and Gibbs sampling (Geman and Geman, 1984) algorithms are traditional MCMC sampling algorithms that simulate a Markov chain to sample from a given distribution. A major challenge associated with such traditional MCMC sampling methods is that the burn-in period for the Markov chain is unknown, and being dependent on the model and the data, difficult to predict in advance. To have a high assurance of convergence of the Markov chain, we need to run the sampler for very long making it computationally expensive. More advanced MCMC methods like CFTP (Propp and Wilson, 1996) and FA (Fill, 1998) provide finite-time algorithms (that can be far more practical than traditional MCMC) to guarantee convergence of the Markov chain, thereby assuring sampling from the true distribution. FA has clear advantages over CFTP that can introduce a bias in the sampling stemming from possible user impatience leading to forced early termination due to computational costs. Nevertheless, CFTP and FA proposed in Propp and Wilson (1996); Fill (1998) are applicable only to a small class of MRF models that lead to *monotonic* chains. Later theoretical advancements propose the concept of the BC (Huber, 2004) to provide efficient ways to track non-monotonic chains to detect Markov-chain convergence. However, the BC algorithm in Huber (2004) is restricted to the CFTP paradigm and can suffer from the user-impatience bias problem. Thus, we propose FA-BC as a novel theoretical extension of FA applicable to general MRF models. None of the aforementioned works (Propp and Wilson, 1996; Fill, 1998; Huber, 2004) apply their theory to the problem of MRF-based image segmentation or uncertainty estimation. Our work is perhaps the first to apply the perfect-sampling frameworks for sampling label images in MRF models and estimate uncertainty.

Some recent theoretical advances propose the perturb-and-MAP framework and a Gumbel perturbation model (GPM) (Papandreou and Yuille, 2011; Hazan et al., 2013) to exactly sample from MRF distributions. However, GPMs are practically intractable for large-sized MRFs, as observed in image segmentations applications. Alberts et al. (2016) employed Gumbel MAP perturbations for a grid-structured MRF for brain tumor segmentation and quantifying uncertainty. In contrast, we introduce the *perfect / exact MCMC* paradigm and propose a novel perfect-MCMC sampler for *generic Bayesian MRFs*, to estimate uncertainty in multilabel and multiatlas segmentation problems.

An alternative to sampling is to approximate the posterior MRF model on the label image with a simpler model, typically a factored model, which is easier to analyse analytically or sample from. Examples of approximate models include those inferred using VB methods (Bishop, 2006), e.g., mean-field approximations and GP models. Unlike MRFs on discrete labels, (Figueiredo, 2005) explores priors on simpler and more tractable models, specifically, real-valued hidden Gaussian field (GF) that indirectly indicate the discrete-label probability, and thereby avoids the intractable combinatorial nature of the problem. However, GFs have quadratic MRF potential functions that cannot model discontinuities in the image data, or crisp object boundaries, effectively. This approach of modeling the spatial prior is closely related to the well-known GP model (Williams and Barber, 1998).

For problems in medical image segmentation addressed in this paper, several approaches have been popular in the literature, although none of these approaches produce estimates of uncertainty in the resulting label image. For tissue segmentation in brain magnetic resonance imaging (MRI), (Zhang et al., 2001) use EM with a Gaussian mixture model (GMM) and a hidden-MRF label prior. However, they use a heuristic to approximate sampling from the posterior in the E step. Pham and Prince (1999) generalize EM algorithm for tissue segmentation using two kinds of MRF priors. Some approaches propose nonparametric (non-local) patch statistical models (Awate et al., 2006a; 2006b), nonparametric models in Riemannian spaces (Awate et al., 2007; Goh et al., 2011), and intensity-Based Markov priors (Song et al., 2007) for image segmentation. While Song et al. (2006) use a graph-cuts based algorithm for brain image segmentation, (Veni et al., 2013) uses constrained graph-cuts to segment nested surfaces in cardiac images. For tumor segmentation, (Fletcher-Heath et al., 2001) propose a fuzzy-clustering based algorithm for separating brain tumors from healthy tissues, (Han et al., 2011) use graph-cut based

segmentation algorithm to find the globally optimal solution, and Shah et al. (2018, 2019) proposes deep neural networks (DNNs) with mixed-supervision learning that combines training data of varying qualities synergistically. For segmenting subcortical structures, (Wolz et al., 2009) propose graph-cut based algorithm and Gouttard et al. (2007) incorporate probabilistic atlas priors to get a maximum-a-posteriori (MAP) solution. For multiatlas segmentation (Iglesias and Sabuncu, 2015) of subcortical structures, (Awate et al., 2012; Awate and Whitaker, 2014) propose bootstrap resampling to learn nonparametric regression models and error-convergence rates indicating voxelwise uncertainty for a population of images (not a specific image). For subcortical structure segmentation, we also propose and evaluate a shape Boltzmann machine (SBM) (Eslami et al., 2014) prior model as a shape prior for binary label images. SBMs use a hierarchical / multilayer structured MRF that consists of a visible layer corresponding to the (unknown) label image, and hidden layers that correspond to latent variables capable of learning higher-order features and inter-voxel correlations capture object shape. Nevertheless, our focus is on uncertainty estimation using various prior models, instead of the segmentation itself. For brain parcellation, (Pohl et al., 2007) propose a algorithm using the hierarchy of anatomical structures and Sabuncu et al. (2009) use an EM-based nonparametric method.

For a small class of MRF models that infer the segmentation using graph cuts, Kohli and Torr (2008) proposes a measure of uncertainty and an efficient algorithm to estimate the same. For general MRFs, typical uncertainty estimation methods approximate the MRF models or use traditional MCMC methods to sample from them. For instance, Fan et al. (2007) uses traditional MCMC to sample nonparametric curves and Beutner et al. (2009) uses traditional MCMC to estimate uncertainty for brain structure delineations. Le et al. (2016) uses a GP approximation for label distributions. In tumor segmentation, Alberts et al. (2016) approximates the GPM in Papandreou and Yuille (2011) and Hazan et al. (2013) to sample from the underlying Bayesian MRF. Unlike traditional MCMC that is only asymptotically exact and can suffer from insufficient burn-in (fixing one very large burn-in for *all* tasks makes computational costs exorbitant), we guarantee exact MCMC in *finite time* and eliminate adhoc heuristics to determine burn-in. A very recent method (Jena and Awate, 2019) estimates uncertainty in medical image segmentation produced by DNNs using a Bayesian formulation that treats the output of the DNN as a per-voxel factored distribution.

This work significantly extends our preliminary work in Garg and Awate (2018). This paper provides significantly more details of the theoretical analysis of sampling, modeling, as well as inference schemes. This paper includes validation against specially designed simulated data for which we have computed the ground truth exactly. This paper demonstrates the applicability of our approach to more MRF models, including GF and SBM, in addition to Ising and Potts MRF models. This paper includes comparisons against posterior-model-approximation schemes underlying VB inference.

3. Theory

We introduce our framework for perfect MCMC sampling to estimate uncertainty in image segmentation. Within the framework, we propose two algorithms, i.e., CFTP-BC and FA-BC. FA-BC has clear advantages over CFTP-BC in theory and practice and, thus, is our algorithm of choice.

3.1. Background

This section describes the background theory and the associated notation for MCMC sampling and perfect sampling.

3.1.1. MCMC sampling

Let the observed image y, with V voxels, be generated from (i) a hidden label image x that is modeled by MRF X with prior probability mass function (PMF) $P(X)$ and (ii) a likelihood model $P(Y|X)$. MRF X has a neighborhood system $\mathcal{N} := \{\mathcal{N}_v\}_{v=1}^{V}$, where $\mathcal{N}_v$ is the set of voxels neighboring voxel v. The posterior PMF $Q(X) := P(X|y)$ has a normalizing constant that is intractable for large-sized images. Thus, to sample from the posterior PMF $Q(X)$, MCMC methods carefully design a Markov chain $\mathcal{M}$ as the MRF sequence X^1, X^2, $\cdots$, X^t, $\cdots$ with an associated *transition kernel* $K(\cdot, \cdot)$ such that (i) the transition probabilities are $P(X^{t+1}|X^t) := K(X^t, \cdot)$ and (ii) the Markov chain's *stationary* PMF is $Q(X)$. Typically, the Markov chain $\mathcal{M}$ is positive, recurrent, and aperiodic. Such a chain is called *ergodic* and has a unique stationary PMF that equals $Q(X)$. For a recurrent chain, the average number of visits to any state (in our case, label-image instance) is ∞. Typically, the Markov chain $\mathcal{M}$ also satisfies *detailed balance*, or *reversibility*, which implies that kernel $K(\cdot, \cdot)$ also applies to the time-reversed chain. The M–H algorithm is a MCMC method for sampling from a high-dimensional PMF having an intractable normalizing constant. The Gibbs sampler is one such M–H MCMC sampler, and has an associated ergodic reversible Markov chain. Gibbs sampling for images iteratively selects a random voxel and samples a value for that voxel from its local conditional PMF. The Gibbs sampler, like typical M-H samplers, requires the Markov chain to be run *infinitely* long to theoretically guarantee that the resulting state will be a draw from the PMF $Q(X)$. In practical scenarios, the convergence time for typical M–H samplers is very difficult to predict, being dependent on the model, model parameters, and the data instance. This poses challenges in being able to apply them seamlessly across a wide spectrum of practical scenarios.

3.1.2. CFTP for perfect MCMC in monotone-chain models

For discrete random variables, a major breakthrough in being able to detect MCMC-sampler convergence came through the CFTP algorithm published in the statistical mechanics literature (Propp and Wilson, 1996). CFTP theoretically guarantees the sampled state to be from the desired PMF $Q(X)$ by ensuring that any long-running Markov chain, irrespective of its initial state, would have reached the chosen sampled state, using a specific sequence of interstate-transition maps.

For image segmentation applications, especially uncertainty estimation, we introduce the perfect-sampling paradigm, starting with the CFTP framework on this section, to sample from MRF posteriors. Later sections describe an even better perfect-MCMC framework.

CFTP tracks *coupled parallel* chains, i.e., (i) one chain started in each possible state of the state space and (ii) all chains using the same pseudo-random number sequence for transitioning. The tracking continues until all of them reach, or *coalesce* to, a single state.

Theorem 1. *Propp and Wilson (1996): The CFTP algorithm terminates in finite time and returns a draw from the stationary distribution of the Markov chain.*

We interpret this theorem as follows. Markov chain ergodicity implies that, for all states x, there is a non-zero probability $> \epsilon > 0$ of reaching x, from any state x' in âa finite number of transitions N_x. For a given instance of a sequence of interstate-transition maps (or, equivalently, random numbers) in the Markov chain, coalescence to some state must occur for some finite number of transitions $M \geq \max_x N_x$. Indeed, the probability of coalescence failing to occur $\rightarrow 0$ as $M \rightarrow \infty$. M is almost-surely *finite* because the probability of coalescence in any finite number of transitions is positive. In practice, for all the applications in this paper, we found that coalescence always happened, even though the number of transitions

varied from several tens to several thousands. Assume that coalescence occurred when the chain ran from time $t = -M$ to $t = 0$, using a specific sequence of transition maps. A chain running from $-\infty$ to 0 that uses this sequence of transition maps within $[-M, 0]$ reaches the same state at $t = 0$. Because the state reached by a chain running infinitely long is a draw from the stationary PMF $Q(X)$, the coalesced state at $t = 0$ is a draw from $Q(X)$.

For some PMFs $Q(X)$, the Gibbs sampler is *monotonic* (Propp and Wilson, 1996), i.e., where transitions of coupled chains preserve a *partial order* on the states. Monotonicity allows CFTP to simplify parallel-chain tracking to tracking only two chains, each started from one of the extremal states (minimum and maximum) under the partial order. While monotonicity holds for the special case of the ferromagnetic Ising model, it fails to apply to many popular binary-MRF and Potts-MRF posterior PMFs. For general cases, perfect sampling can use the *bounding chain* principle (Huber, 2004) as we describe next.

3.1.3. CFTP with BC (CFTP-BC) for perfect MCMC in generic MRF models

For Gibbs sampling, CFTP-BC uses the following modified sampler $\mathcal{G}$ to draw label X_v, at each voxel v, from the conditional PMF $P(X_v|x_{-v})$ conditioned on all other label values x_{-v}. For MRF X, $P(X_v|x_{-v}) = P(X_v|x_{\mathcal{N}_v})$.

1. Draw label l uniformly from the label set $\mathcal{L} := \{1, \cdots, L\}$. Draw $u \sim U(0, 1)$, where $U(a, b)$ is a uniform distribution over (a, b).
2. If $u \leq P(X_v = l|x_{-v})$, set $X_v := l$ and terminate; otherwise, iterate.

Provably, $\forall l$, the probability that this *modified Gibbs sampler* $\mathcal{G}$ terminates with label $X_v = l$ is $P(X_v = l|x_{-v})$, as desired.

For coupled parallel chains, each running a coupled modified Gibbs sampler $\mathcal{G}$, the BC algorithm (Huber, 2004) efficiently tracks the states of all chains. The BC algorithm works for both monotone and non-monotone chains. CFTP-BC uses this tracking strategy to detect coalescence. Consider a new kind of a Markov chain $\mathring{\mathcal{M}}$ with state space $(2^{\mathcal{L}})^V$, where $2^{\mathcal{L}}$ is the set of subsets of $\mathcal{L}$. For the new Markov chain $\mathring{\mathcal{M}}$, each state, say, $\mathring{X}$, contains a set of states X from the earlier Markov chain $\mathcal{M}$, where each state $X \in \mathcal{L}^V$. $\mathring{\mathcal{M}}$ is associated with a state sequence $\mathring{X}^1, \mathring{X}^2, \cdots$ where the transition kernel $\mathring{K}(\cdot, \cdot)$ on $\mathring{X}$ is defined in terms of the transition kernel $K(\cdot, \cdot)$ acting on each state $X \in \mathring{X}$.

Definition 1. Huber (2004): $\mathring{\mathcal{M}}$ is a bounding chain for $\mathcal{M}$ if there exists a coupling between $\mathring{\mathcal{M}}$ and $\mathcal{M}$ such that $X_v^t \in \mathring{X}_v^t, \forall v, \Rightarrow X_v^{t+1} \in \mathring{X}_v^{t+1}, \forall v$.

Consider all coupled parallel chains $\mathcal{M}$ running $\mathcal{G}$ and visiting voxel v at time t. The bounding chain $\mathring{\mathcal{M}}$ keeps track of the set $\mathring{X}_v \subseteq \mathcal{L}$ of possible labels, at each v, across all chains $\mathcal{M}$ at any given time; it initializes $\mathring{X}_v := \mathcal{L}$ and detects coalescence when $|\mathring{X}_v| = 1, \forall v$. Each chain $\mathcal{M}$ has its conditional PMFs $P(X_v|x_{-v})$, dependent on MRF-neighborhood configurations $x_{\mathcal{N}_v}$. At any point in time in the progression of the coupled parallel Markov chains $\mathcal{M}$, for each label l, let the minimum and maximum of conditional probabilities $P(X_v = l|x_{-v})$, over all chains $\mathcal{M}$, be $P^{\min}(X_v = l|x_{-v})$ and $P^{\max}(X_v = l|x_{-v})$. Here, the probabilities are considered over all possible neighborhood label configurations $x_{\mathcal{N}_v}$ in the cross-product space $\mathring{X}_{w_1} \times \mathring{X}_{w_2} \times \cdots$, where $w_i \in \mathcal{N}_v$, and capture all possible MRF-neighborhood label configurations $x_{\mathcal{N}_v}$ around voxel v across all chains. We can partition the set of all chains $\mathcal{M}$ into equivalence classes, based on possible MRF-neighborhood label values $x_{\mathcal{N}_v}$, within which the coupled Gibbs samplers $\mathcal{G}$ behave identically at voxel v. Now do the following at voxel v:

1. In the bounding chain $\mathring{\mathcal{M}}$, initialize the set of possible labels $\mathring{X}_v := \emptyset$.
2. Draw l uniformly from the label set $\mathcal{L}$. Draw $u \sim U(0, 1)$.
3. If $u \geq P^{\max}(X_v = l|x_{-v})$, then *no* coupled chain $\mathcal{M}$ has changed state. So, do nothing.
4. If $u \in (P^{\min}(X_v = l|x_{-v}), P^{\max}(X_v = l|x_{-v}))$, then some of the equivalence classes of coupled chains $\mathcal{M}$ have set $X_v \leftarrow l$. So, insert label l into set $\mathring{X}_v$.
5. If $u \leq P^{\min}(X_v = l|x_{-v})$, then all coupled chains $\mathcal{M}$ set $X_v \leftarrow l$, indicating, what we call as, "local" coalescence that is a sufficient condition for every chain $\mathcal{M}$ to have undergone *at least one* transition where its sampler $\mathcal{G}$ terminated. So, insert label l into the set $\mathring{X}_v$. **Exit**.
   The mechanism of the BC $\mathring{\mathcal{M}}$ is cleverly designed to avoid explicitly tracking each chain or each equivalence class, when there are a possibly exponential number of each kind (exponential in the number of image voxels). On the other hand, when the exit criterion is met after some chains $\mathcal{M}$ running $\mathcal{G}$ multiple times, i.e., with multiple terminations, then the BC mechanism includes *all* (intermediate and final) sampled labels in $\mathring{X}_v$, thereby allowing a possibly looser bound that inflates the cardinality of the set $\mathring{X}_v$ of possible label values. Nevertheless, in practice, we find that the chain approaches coalescence rapidly despite this looser bound.
6. Repeat from Step 2.

When, the sets $\mathring{X}_v$ become singletons $\forall v$, say, $\{\widehat{x}_v\}$, then all chains $\mathcal{M}$ have coalesced to label image $\widehat{x}$ that is guaranteed to be a draw from the stationary PMF $Q(X)$. Ergodicity of $\mathcal{M}$ ensures coalescence in finite time, almost surely.

3.1.4. FA for perfect MCMC in monotone-chain models

A limitation of the CFTP strategy proposed in Propp and Wilson (1996), exhibited by the monotone-chain CFTP algorithm in Propp and Wilson (1996) as well as the CFTP-BC algorithm in Huber (2004), is that the CFTP running time M and the sampled state $\widehat{X}$ are *dependent* variables. Even when M is finite, M is *unbounded* whose order of magnitude is typically unknown a priori; M depends on several factors including the pseudo-random sequence, the posterior model and its parameter values, and the data. Thus, some states x can require a very long run from $-M$ to 0, with large unpredictable M. Impatient users abort CFTP when M starts becoming large, which adds bias to the sampled states' PMF. In contrast, FA (Fill, 1998) makes the sampled state independent of FA's running time. FA relies on acceptance-rejection (AR) sampling that first proposes a state and then decides whether to reject the proposal. The FA described in Fill (1998) works *only for monotone* $\mathcal{M}$, as below.

1. Choose T in increasing progression as 1, 2, 4, 8, $\cdots$. For each T in this sequence, perform the following steps, until termination.
2. Choose a random label image $X^T := z$, uniformly from the space of label images.
3. Run a (time-reversed) Markov chain $\mathcal{M}$ from time $t = T$ down to time $t = 0$, with initial $X^T := z$, reaching some new state $X^0 := x$. Note: One could treat this as a forward chain, and then treat the chain mentioned in the next step as the time-reversed chain.
4. Let $S^T(x, z)$ be the event that a Markov chain starting at state x at time $t = 0$ ran for time T to reach state z at time $t = T$. The event $S^T(x, z)$ occurs for some set of pseudo-random number sequences $\mathcal{U}^{x \to z}$. Let $C^T(z)$ be the event that the set of coupled parallel chains, one starting at every state in the state space, ran for time T and coalesced in z; this occurs for some set of pseudo-random number sequences $\mathcal{U}' \subseteq \mathcal{U}^{x \to z}$. With probability $P(C^T(z)|S^T(x \to z))$, accept x as a draw from the stationary PMF $Q(X)$ and terminate; otherwise iterate from Step (2) with increased T.

For the case of monotonic chains, Step 4 can be simulated as follows. Begin two coupled Markov chains at the extremal states and run them for T steps. If the two chains coalesce in state z, then accept x as a draw from the stationary distribution.

Theorem 2. *Fill (1998): Fill's algorithm, with constrained monotone chains, guarantees that the sampled state is from the stationary PMF $Q(X)$.*

This is true because of the following arguments. The AR sampler underlying FA first generates a proposal x from the T-step transition kernel $K^T(z, \cdot)$. The reversibility of the Markov chain implies that $Q(y)K^T(y,z) = Q(z)K^T(z,y)$, for any states y, z. Fixing z, we have $Q(y) = Q(z)K^T(z,y)/K^T(y,z)$ that holds for all states y. Because we have a discrete state space, $K^T(y,z) \geq P(C^T(z))$, $\forall y$. Thus, $Q(y) \leq Q(z)K^T(z,y)/P(C^T(z))$, $\forall y$. Thus, $K^T(z,\cdot)Q(z)/P(C^T(z))$ is a function that upper bounds the stationary PMF $Q(\cdot)$. If we define $M_z^T := Q(z)/P(C^T(z))$, then the upper-bounding function is $M_z^T K^T(z,\cdot)$. The AR sampler then accepts the proposal x with a probability equal to $Q(x)/(M_z^T K^T(z,x)) = Q(x)P(C^T(z)/(Q(z)K^T(z,x)) = Q(x)P(C^T(z)/(Q(x)K^T(x,z))$, due to reversibility, which in turn equals $P(C^T(z)|S^T(x,z))$. The AR sampler follows the following strategy. In a single iteration, (i) the probability of generating a state x is $K^T(z,x)$ and (ii) the probability of accepting a given x is $Q(x)/M_z^T K^T(z,x)$. Thus, the probability of generating a state x and accepting it, in a single iteration, is $Q(x)/M_z^T$. Also, the unconditional acceptance probability of the AR sampler is $1/M_z^T$. Thus, conditioned on the acceptance at this iteration, the acceptance probability of x is $Q(x)$, as desired.

3.2. Proposed FA with BC (FA-BC) for perfect MCMC in generic MRF models

Previous works limit Fill's algorithm to monotone chains that apply to a very small class of PMFs $Q(X)$. For monotone chains, detecting $C^T(z)$ constrained on $S^T(x \to z)$ needs the tracking of only two extremal states. Even though evaluating $P(C^T(z)|S^T(x \to z))$ is computationally intractable, for general cases, AR decisions can be made by (i) drawing a $\sqcap^{x \to z} \in \mathcal{U}^{x \to z}$, drawn uniformly randomly, to ensure $S^T(x \to z)$ occurs and (ii) tracking coupled parallel chains, transitioning as per $\sqcap^{x \to z}$, to detect if $C^T(z)$ occurs. We generalize Fill's algorithm to generic Bayesian MRFs by efficiently tracking constrained parallel *arbitrary* chains using a novel *constrained-BC* algorithm, as follows.

At time t and voxel v, for each label l, let $P^{\min}(X_v^t = l|x_{-v}^t)$ and $P^{\max}(X_v^t = l|x_{-v}^t)$ be defined as before. Let l^* be the label at voxel v for time $t+1$ along some T-step Markov chain path $x \to z$. While this path may be chosen at random from the set of all T-step paths starting at x and ending at z, in practice, we should explore close to, and starting with, the reverse sequence of the observed path sequence from $z \to x$ that generated the AR sampler's proposal x. The observed path, and nearby paths, can lead to a better choice than randomly drawing $x \to z$ paths, because the latter may lead to very-low-probability unrealistic paths that fail to lead to coalescence of the FA-BC sampler. Indeed, the larger the T, the more it is likely that the state x can be taken as a draw from the desired stationary PMF. In this paper, we only use the reverse of the path $z \to x$ to draw a random sequence $\sqcap^{x \to z}$ from $\mathcal{U}^{x \to z}$. At time t, let $P^*(X_v^t = l^*|x_{-v}^t)$ be the label probability conditioned on neighboring labels for the chosen path $x \to z$. Clearly, $P^{\min}(X_v^t = l|x_{-v}^t) \leq P^*(X_v^t = l^*|x_{-v}^t) \leq P^{\max}(X_v^t = l|x_{-v}^t)$. Initialize $t := 0$, $x^0 := x$.

1. At time t, do the following at each voxel v:
   (a) In the bounding chain $\mathring{\mathcal{M}}$, initialize the set of possible labels $\mathring{X}_v := \emptyset$.
   (b) Draw l uniformly from the label set $\mathcal{L}$.
   (c) If $l \neq l^*$, draw $u \sim U(P^*(X_v^t = l|x_{-v}^t), 1)$; otherwise draw $u \sim U(0,1)$. This sampling strategy leads to a random draw $\sqcap^{x \to z} \sim \mathcal{U}^{x \to z}$, ensuring that x^t transitions to x^{t+1} on the path from $x \to z$, thereby leading to $S^T(x \to z)$. The next steps track parallel coupled chains to detect if $C^T(z)$ occurs for $\sqcap^{x \to z}$.
   (d) If $u \geq P^{\max}(X_v^t = l|x_{-v}^t)$, then *no* chain $\mathcal{M}$ changes state. Go to Step 1b.
   (e) If $u \in (P^{\min}(X_v^t = l|x_{-v}^t), P^{\max}(X_v^t = l|x_{-v}^t))$, then some chains $\mathcal{M}$ accept label l. Insert l into $\mathring{X}_v$. Go to Step 1b.
   (f) If $u \leq P^{\min}(X_v^t = l|x_{-v}^t)$, then all chains $\mathcal{M}$ set $X_v = l$. This is, what we call, "local" coalescence. Insert l into $\mathring{X}_v$. If any voxel remains to be processed, go to Step 1a to process it; otherwise go to the next step.
2. Increment t by 1. If $t < T$, repeat Step 1. If $t = T$ and coalescence has occurred, i.e., $|\mathring{X}_v| = 1, \forall v$, then accept the initial x as a draw from $Q(X)$.

Theorem 3. *Our modification, i.e., FA-BC, of the FA described in Fill (1998) adapted to the constrained bounding chain, guarantees that the sampled state is from the stationary PMF $Q(X)$.*

Proof. We show that our random number generation scheme in Step 1c ensures $S^T(x \to z)$ by simulating a $\sqcap^{x \to z} \in \mathcal{U}^{x \to z}$. At time t and voxel v, let E^* be the event that, for the chain going from $x \to z$, the label at voxel v at time $t+1$ is l^*. Let the $x \to z$ chain's *unconstrained* modified Gibbs sampler be $\mathcal{G}^*$ and the label probabilities be $P^*(X_v^t = l|x_{-v}^t)$. For E^* to occur, $\mathcal{G}^*$ accepted label l^* in some iteration i. In any iteration, $\mathcal{G}^*$ picked some l and some random u. Given E^*: if $\mathcal{G}^*$ picked an $l \neq l^*$, then u must have been within $(P^*(X_v^t = l|x_{-v}^t), 1)$; otherwise u could have been anywhere within (0,1). Now consider *parallel coupled* chains, one starting at each possible state, running sampler $\mathcal{G}$ for T transition steps. At iteration i, if $\mathcal{G}$ picks $l \neq l^*$, then $\mathcal{G}$ must pick u within $(P^*(X_v^t = l|x_{-v}^t), 1)$ because, otherwise, the chain started at x can incorrectly accept $l \neq l^*$ and E^* can fail to occur. At iteration i, if $\mathcal{G}$ picks $l = l^*$, then $\mathcal{G}$ can pick u within (0,1), leading to a finite non-zero probability for the chain started at x accepting l^* and leading to E^*. Steps d-f track all chains, as in CFTP-BC, to detect $C^T(z)$ for the chosen $\sqcap^{x \to z}$. Following the AR-sampler based arguments in the interpretation of Theorem 2 in the earlier section, we can guarantee that the sampled state is from the stationary PMF. □

4. Applications: uncertainty estimation in image segmentation in various MRF modeling and inference frameworks

We describe the applications in medical image segmentation to which we apply the proposed perfect-sampling methods, along with contemporary approaches, to estimate uncertainty. For each application, we describe the underlying statistical modeling and inference approaches.

4.1. Applications

We evaluate several methods on four classic problems in medical image segmentation involving in brain MRI images, as detailed in Section 4.1, for segmenting subcortical structures, tumor, tissues, and lobes. (i) EM segmentation of brain tumor, with a hidden-MRF label prior and a two-component GMM, one component each for the tumor and non-tumor intensity patches on multimodal MRI; (ii) EM segmentation of brain tissues, involving mild lesions, with a hidden-MRF label prior and a three-component GMM, one component each for gray matter (GM) intensity, white matter (WM) intensity and cerebrospinal fluid (CSF) intensity patches; (iii) multiatlas segmentation of subcortical brain structures; and (iv) multiatlas segmentation of the four brain lobes.

4.2. Data likelihood models, given segmentation label images

For the two EM-based applications, the EM algorithm of Zhang et al. (2001) iteratively estimates the parameters for the GMM. Let the estimated parameters be the Gaussian means $\boldsymbol{\mu} := \{\mu^1, \mu^2, \cdots, \mu^K\}$ and covariances $\boldsymbol{\Sigma} := \{\Sigma^1, \Sigma^2, \cdots, \Sigma^K\}$, where K is number of classes. The likelihood of the intensity patch $x_{\mathcal{N}_i}$ at voxel i being drawn from class label $y_i = k$ is

$$P(x_{\mathcal{N}_i}|y_i = k; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\Sigma^k|^{-0.5} \exp\Big(-0.5(x_{\mathcal{N}_i} - \mu^k)^\top (\Sigma^k)^{-1}(x_{\mathcal{N}_i} - \mu^k)\Big). \tag{1}$$

The multiatlas segmentation applications use the classic voxel-wise nonparametric label-likelihood model (Iglesias and Sabuncu, 2015; Awate and Whitaker, 2014), as follows. Let the multiatlas database $\mathcal{D} := \{z^j, s^j\}_{j=1}^J$ have template MRI images z^j paired with label images s^j. At voxel i, the observed-image patch $x_{\mathcal{N}_i}$ has likelihood

$$P(x_{\mathcal{N}_i}|Y_i = l, \mathcal{D}) := \sum_{j=1}^J \mathbf{1}_l(s_i^j) G(\check{x}_{\mathcal{N}_i}; \check{z}^j_{\mathcal{N}_i}, \sigma^2\mathbf{I}) \Big/ \sum_{j=1}^J \mathbf{1}_l\big(s_i^j\big), \tag{2}$$

where $\mathbf{1}_l(a) := 1$ if $l = a$ and 0 otherwise, $\mathbf{I}$ is the identity matrix, σ^2 the Gaussian kernel variance, and $\check{x}_{\mathcal{N}_i}$ and $\check{s}^j_{\mathcal{N}_i}$ are normalized patches with mean 0 and variance 1.

4.3. Prior models on segmentation label images

Our applications employ three different probabilistic prior models on label images, i.e., the MRF, the GF, and the SBM.

4.3.1. Ising/potts MRF prior model

For voxel i, the prior conditional probability of label $y_i = k$, given neighborhood labels, is

$$P(y_i = k|\mathbf{y}_{\mathcal{N}_i}) \propto \exp\Big(\sum_{j\in\mathcal{N}_i} w_{ij}\mathbf{1}_k(y_j)\Big), \tag{3}$$

where w_{ij} is the weight (interaction) parameters pre-defined for the MRF model.

4.3.2. GF-based MRF prior model

Unlike the MRF model on discrete labels, we explore a prior model on real-valued hidden fields (Figueiredo, 2005) where the real values indicate label probabilities. Consider a collection of real-valued hidden fields $\mathbf{z} = \{z^1, z^2, \cdots, z^K\}$ that indicate the label probabilities for K classes, at each voxel. Spatial regularization on these real-valued fields leads to spatially smooth label probabilities. This approach is akin to the statistical modeling technique of Gaussian processes and Gauss MRFs. The label probability at voxel i is given by a multinomial logistic model (softmax) as

$$P(y_i = k|\mathbf{z_i}) = \exp(z_i^k) \Big/ \sum_{k=1}^K \exp\big(z_i^k\big). \tag{4}$$

The set of hidden fields are independent, i.e., $P(\mathbf{z}) = \prod_{k=1}^K P(z^k)$, and each field's probability model is the Gauss MRF

$$P(z^k) \propto \exp\Big(-\sum_i \sum_{j\in\mathcal{N}_i} w_{i,j}\big(z_j^k - z_i^k\big)^2\Big), \tag{5}$$

where $w_{i,j}$ are pre-defined neighborhood weighting parameters. The GF model is simpler and more tractable than the MRF model on discrete label images, but has limited ability to model discontinuities in the image data.

4.3.3. SBM-based MRF prior model

The SBM model relies on concepts related to the restricted Boltzmann machine (RBM) (Salakhutdinov et al., 2007) and the deep Boltzmann machine (DBM) (Salakhutdinov and Larochelle, 2010). So, we first describe the RBM and the DBM, and then the SBM. Note: we use the SBM prior only for multiatlas segmentation of subcortical structures to capture the shape variability in the structures. Typical MRF models are good at learning local conditional dependencies due to the nature of the graph. However, typical MRF models can fail to model long-range dependencies effectively. For instance, object pose is a "global" feature that is difficult to capture using a MRF with small neighbourhoods. An effective way of modeling long-range dependencies between random variables is by using a multi-layer MRF model. One such MRF model is the RBM.

A RBM partitions the underlying random variables X into (i) "visible" units $V = \{V_1, V_2, \cdots, V_m\}$ that are the label variables per voxel, and (ii) "hidden" units $H = \{H_1, H_2, \cdots, H_n\}$ that capture long-range dependencies. The visible and hidden units take values binary values (0 or 1), and the joint model probability is given by the Gibbs PMF $P(V, H) := (1/\eta)\exp(-E(V, H))$, where the Gibbs energy $E(V, H) := -V^\top WH - b^\top V - c^\top H$, where W is a weight matrix of size $m \times n$, b is a bias vector for visible units of size $m \times 1$, c is a bias vector for hidden units of size $n \times 1$, and η is the partition function. The form of the Gibbs energy implies MRF-neighborhood connections between visible and hidden layers, but *not* among units of the same layer. Therefore, the visible units interact indirectly through the hidden layer units, and because all visible units are connected to any hidden unit (and vice versa), all the units can interact with each other indirectly. The Gibbs energy also implies that the visible units are independent given the hidden units, and vice versa. Thus, $P(V|H) = \prod_{i=1}^m P(V_i|H)$ and $P(H|V) = \prod_{j=1}^n P(H_j|V)$. This factorization enables to efficiently use block Gibbs sampling. Also, the Gibbs energy leads to the following equations (where $\mathcal{S}(\cdot)$ is the sigmoid function): $P(V_i = 1|H) = \mathcal{S}(W_i H + b_i)$ and $P(H_j = 1|V) = \mathcal{S}(W_j^\top V + c_j)$. RBM learning uses the contrastive divergence (Bishop, 2006) method to approximate maximizing the data likelihood. The limitations of the RBM arise from having a single layer of hidden units, which limits its modeling capacity. Also, there are no constraints on local and global spaces. These problems are alleviated by the DBM.

DBM allows us to efficiently model complex long-range dependencies between the visible units (voxel labels) with far fewer number of units than an equivalent RBM. A DBM comprises at least two hidden layers. For the simplest form of DBM, let the two hidden layers be $H^{(1)}$, and $H^{(2)}$. Then, the probability model is $P(V, H^1, H^2) = (1/\eta)\exp(-E(V, H^1, H^2))$, where $E(V, H^1, H^2) = -V^\top W^1 H^1 - (H^1)^\top W^2 H^2 - b^T V - (c^1)^\top H^1 - (c^2)^\top H^2$. Although the multi-layer model makes DBMs more powerful than RBMs, but this comes at the cost of learning a very large number of parameters. On the other hand, the SBM helps regularize the DBM through appropriate constraints in the context of image data.

The main difference between a DBM and a SBM is that the SBM divides the visible layer into patches and the first layer of hidden units are different for different patches. That reduces the number of parameters and encourages the network to learn local patch structure. The second layer of hidden units is connected to all the first hidden layer units. For example, if an image is broken into four patches V^1, V^2, V^3, V^4, there would be four patches of first layer of hidden units H^{11}, H^{12}, H^{13}, H^{14}, and the second layer of hidden units is H^2 as before. SBM's energy function is $E(V, H^1, H^2) = \sum_{i=1}^4 \big[-V^i W^{1i} H^{1i} - H^{1i} W^{2i} H^{2i} - (c^{1i})^\top H^{1i} - (b^i)^\top V^i\big] - (c^2)^\top H^2$. SBM training and inference follows variational inference methods. Initializing the SBM-layer parameters by pre-training consecutive pairs of layers as RBMs can make the SBM

learning faster and more robust against local minima. For segmentation of a subcortical brain structure, we learn a SBM prior model using binary ground-truth segmentations of that structure. Unlike the GF prior, the SBM prior for multiatlas segmentation makes it easy to calculate the exact marginals for the posterior distribution, once we solve for the optimal values of the hidden layers. This is because, given the hidden-layer values, the prior model gets factorized (as per 4) just like the likelihood models (as per 1 and 2).

4.4. Uncertainty estimation in image segmentation

Our goal is to compute the uncertainty in segmentation as the variance of the per-voxel marginal for the posterior distribution on the label image. For each application, i.e., EM or multiatlas segmentation, we first fit the posterior to the data, optimizing the underlying parameters to maximize the posterior, and then evaluate the marginal variances underlying the posterior. For some posterior models, when the posterior marginals cannot be evaluated analytically, we employ approximation based schemes, to evaluate the posterior mean and variances, based on (i) fitting a factorized PMF to the true posterior using VB inference, or (ii) sampling, i.e., Monte Carlo approximation. Sampling approaches are further categorized into non-perfect sampling and the proposed perfect sampling.

4.4.1. Model approximation

An alternative to sampling is to approximate the posterior by a factorized distribution. Consider the label-image posterior $P(Y|X)$ and an approximated factorized model $Q(Y)$ corresponding to the partition of elements of Y into disjoint groups Y_i, for $i = 1, 2, \cdots, N$, such that

$$Q(Y) := \prod_{i=1}^{N} Q_i(Y_i). \tag{6}$$

From the family of distributions modeled by $Q(Y)$, the VB approach (Bishop, 2006) selects the distribution that minimizes the Kullback–Leibler (KL) divergence from $Q(Y)$ to $P(Y|X)$, which is equivalent to maximizing the log-posterior-function lower bound

$$\mathcal{L}(Q) = \sum_{Y} Q(Y) \log \frac{P(X,Y)}{Q(Y)}. \tag{7}$$

For notation simplicity we refer $Q_i(Y_i)$ as Q_i. Substituting (6) in (7) gives

$$\mathcal{L}(Q) = \sum_{Y} \prod_{i} Q_i \Big(\log P(X,Y) - \sum_{i} \log Q_i \Big). \tag{8}$$

To optimize the jth factor Q_j, we gather terms containing Q_j to give

$$\begin{aligned} \mathcal{L}(Q) &= \sum_{Y_j} Q_j \sum_{Y_i : i \neq j} \Big(\log P(X,Y) \prod_{i \neq j} Q_i \Big) - \sum_{Y_j} Q_j \log Q_j + \text{constant} \\ &= \sum_{j} Q_j \log \widetilde{P}(X, Y_j) - \sum_{Y_j} Q_j \log Q_j + \text{constant}, \end{aligned} \tag{9}$$

where

$$\widetilde{P}(X, Y_j) = \mathbb{E}_{i \neq j}[P(X,Y)] + \text{constant}. \tag{10}$$

If we keep all factors $Q_{i:\, i \neq j}$ fixed and maximize $\mathcal{L}(Q)$ in (9) with respect to all possible forms of Q_j, then the optimal Q_j is

$$\log Q_j^* = \mathbb{E}_{i \neq j}[P(X,Y)] + \text{constant}. \tag{11}$$

The additive constant in (11) gets absorbed in the normalizing constant, such that

$$Q_j^* = \frac{\exp(\mathbb{E}_{i \neq j}[P(X,Y)])}{\sum_{Y_j} \exp(\mathbb{E}_{i \neq j}[P(X,Y)])}. \tag{12}$$

Thus, the VB approach finds the optimal factors by first initializing all of the factors Q_j and then cycling through the factors. We use this factored model Q, approximating the true posterior, to estimate the marginals for each voxel and to obtain the uncertainty.

4.4.2. Sampling (Non-Perfect)

We evaluate (i) classic Gibbs sampling (Geman and Geman, 1984) and (ii) the recent aGPM (Alberts et al., 2016) for MRF posterior models on label images. The burn-in for a Gibbs sampler is notoriously difficult to predict, and depends on the specific MRF model and the dataset. On one hand, choosing a conservative burn-in to maintain the accuracy of the sampler leads to exorbitant computational cost. On the other hand, choosing a small burn-in fails to guarantee samples from the true distribution. Unlike sampling methods which simulate Markov chain to sample from the stationary distribution, Gumbel perturbation sampling methods (Hazan et al., 2013; Papandreou and Yuille, 2011) draw probably-approximate samples from the underlying Gibbs distribution by solving a MAP assignment problem in the perturbed distributions. Let $\{\gamma(y)\}_{y \in \mathcal{Y}}$ be a collection of Gumbel random variables with zero mean and variance c (Euler-Mascheroni constant) associated with label images y in the label-image space $\mathcal{Y}$. Consider random MRF-potential functions of the form $E(y) + \gamma(y)$, where $E(y)$ is the Gibbs energy associated with the underlying sampling distribution and random i.i.d. perturbations $\gamma(y)$ are applied to each label image $y \in \mathcal{Y}$. Then, Theorem 1 in Hazan et al. (2013) states that $(1/Z)\exp(E(\hat{y})) = P_\gamma\big(\hat{y} \in \arg\max_{y \in Y}(E(y) + \gamma(y))\big)$, i.e. solving the MAP assignment problem in a Gumbel-perturbed Gibbs-energy landscape is equivalent to sampling from the MRF associated with the Gibbs energy. However, for sampling segmentations in real-world images, this is practically infeasible because the state space is huge. Thus, the recent method in Alberts et al. (2016) approximates the perturbations by perturbing unary potentials of the underlying graphical model with zero-mean independent an identically-distributed Gumbel random variates.

4.4.3. Proposed perfect sampling methods to estimate uncertainty in segmentation

We propose to estimate uncertainty in MRF-based image segmentation using perfect sampling from the true MRF-based posterior distribution on the label-image. We propose perfect sampling using CFTP-BC and our novel FA-BC samplers. However, we find that the advantages of CFTP-BC are subsumed within the advantages of FA-BC. Thus, our evaluation mainly restricts to the FA-BC sampler.

We apply our FA-BC perfect-MCMC sampler to estimate uncertainty and compare it with the aforementioned (baseline) approaches that either approximate the model or approximate the sampling. We use FA-BC (i) during parameter estimation via EM, in the E step for Monte Carlo sampling label image X from its posterior, and (ii) after parameter estimation, to estimate uncertainty by sampling label maps from the posterior, given optimal parameters, and measuring their variability per voxel.

5. Results and discussion

We show results on simulated data and on four classic brain-MRI analyses, i.e., segmenting subcortical structures, tumor, tissues, and lobes. We evaluate several classes of methods: (A) perfect sampling on the true posterior, i.e., our method using FA-BC from Section 3; (B) approximate sampling on the posterior that includes (B.1) the recent aGPM (Alberts et al., 2016) and (B.2) the traditional Gibbs sampler; (C) approximate factorized modeling of the posterior using the VB approach in Section 4.4; (D) simpler tractable

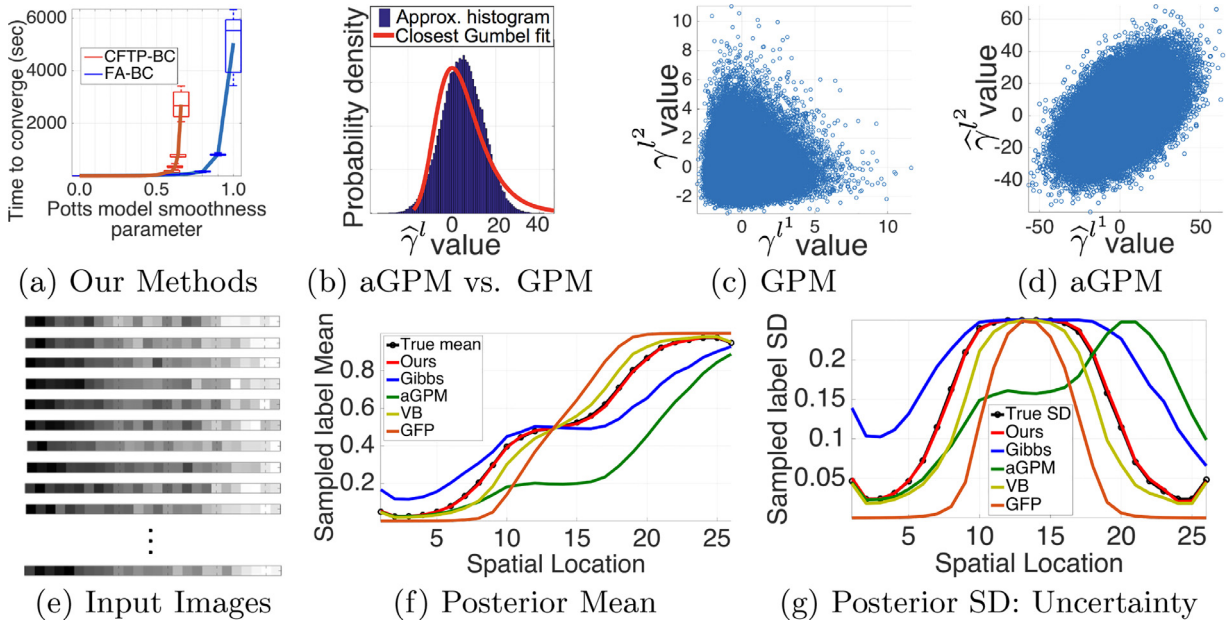

Fig. 1. Validation on simulated data. (a) Our FA-BC and our CFTP-BC: convergence time for sampling label images from a standard Potts MRF model. (b)–(d) Differences between *ideal* Gumbel perturbations γ in Papandreou and Yuille (2011) (intractable for label-image sampling) and their tractable *approximations* $\widehat{\gamma}$ in aGPM (Alberts et al., 2016): (b) For a label image l, empirical histogram for $\widehat{\gamma}^l := \sum_{i=1}^{128} \gamma_i^{l_i}$, as per aGPM's notation, is almost Gaussian (central limit theorem), deviating significantly from Gumbel. (c)–(d) For label images l^1 and l^2, scatter between aGPM draws $\widehat{\gamma}^{l^1}$ and $\widehat{\gamma}^{l^2}$ (both using *same* sample for $\gamma_i^{\star}$) deviates from that between Gumbel draws γ^{l^1} and γ^{l^2}. (e) Simulated 26-voxel 1D input images, each representing the same image edge with a different noise instance. (f)-(g) Label-image-posterior mean and SD (voxelwise) of label images underlying the hidden-MRF posterior for various methods, averaged over the 70 simulated image instances as in (e).

prior model including the GF model in Section 4.3, which has limited ability to capture discontinuities in image data. Methods in classes (C) and (D) work without the need for sampling because, after the simplification, the variation in the label probabilities can be obtained analytically. For the specific application of multiatlas segmentation, we also include results using the SBM model that is a more complex hierarchical MRF prior model on label images, as described in Section 4.3. For the SBM prior, we show results using all methods in classes (A), (B), and (C).

For all sampling-based approaches, we use a large sample size of 200 for the label images; we found that sample sizes larger than 200 left the estimates virtually unchanged. From the sampled set of label images, we compute the mean and variation of the object label per voxel. For the two-class case, we use the standard deviation (SD) as a measure of variation; for the multi-category case, we generalize SD by square-root of unalikeability (Perry and Kader, 2005; Kader and Perry, 2007).

For all sampling methods, we initialize the Markov chain by sampling at random and uniformly from the space of label images. For our method using FA-BC, we choose T in increasing progression as 1, 2, 4, 8, $\cdots$, and so on. We generate one sampled state from a single run of Markov chain. To obtain multiple independent sampled states, we repeat the entire process. For Gibbs sampling, we run the Markov chain for 5 iterations to obtain a sample. For all applications with Ising/Potts MRF prior model, we tune and fix the smoothness prior apriori. Because our focus in this work is *not* on image segmentation, but rather uncertainty estimation, we did *not* focus on fine-tuning and selecting the best possible prior models for specific application tasks. On the other hand, our proposed methods of uncertainty estimation are generic enough to apply to a variety of modeling paradigms.

5.1. Validation on simulated data

The advantage of our FA-BC approach over our CFTP-BC approach is clear from the empirical analysis of the sampler convergence time for sampling label images from a standard Potts MRF model (Fig. 1(a)). As the model parameter (say, β) changes to enforce higher spatial regularity, the convergence of CFTP-BC takes far too many transition steps T and computation times, or virtually fails to terminate (for $\beta > 0.66$), becoming intractable far sooner than those for FA-BC. As described before, CFTP-BC risks biases in sampling arising from user impatience, as described in Section 3. Thus, all further evaluation uses our FA-BC method.

Sampling from the label-image posterior using the GPM scheme in Papandreou and Yuille (2011) is intractable for image sizes in clinical practice. Hence, aGPM (Alberts et al., 2016) introduces approximation to make this tractable. The aGPM approximation in Alberts et al. (2016) leads to perturbations that are actually strongly Gaussian, thereby deviating significantly from the true Gumbel perturbations (Fig. 1(b)), for an Ising model on 128-voxel 1D image. Moreover, the aGPM approximation introduces correlations (Fig. 1(d)) between the perturbations associated with two different label images, unlike GPM (Fig. 1(c)).

We compare methods on binary segmentation task on a simulated 26-voxel 1D image that represents an edge across an object's boundary, i.e., low intensities on left outside the object and high intensities on right inside object. We choose the image small enough to enable the brute-force computation of the normalizing constant of the MRF. This allows us to analytically evaluate the mean (segmentation) of the posterior and the associated per-voxel SD. We corrupt the input image with noise, compute the posterior mean and SD, and repeat this on 70 different noisy image instances. We then compute, for all methods, the average of the posterior mean and SD over all the 70 cases. Our FA-BC based empirical estimates of the mean and SD (Fig. 1(e)–(f)) almost perfectly match the ground truth. All other methods lead to mean and SD estimates that deviate significantly from the ground truth. This clearly demonstrates the potential of our FA-BC in generating estimates of the empirical mean and SD far more accurately than all other methods, which we believe to carry forward in the analyses on clinical brain MRI that we describe next.

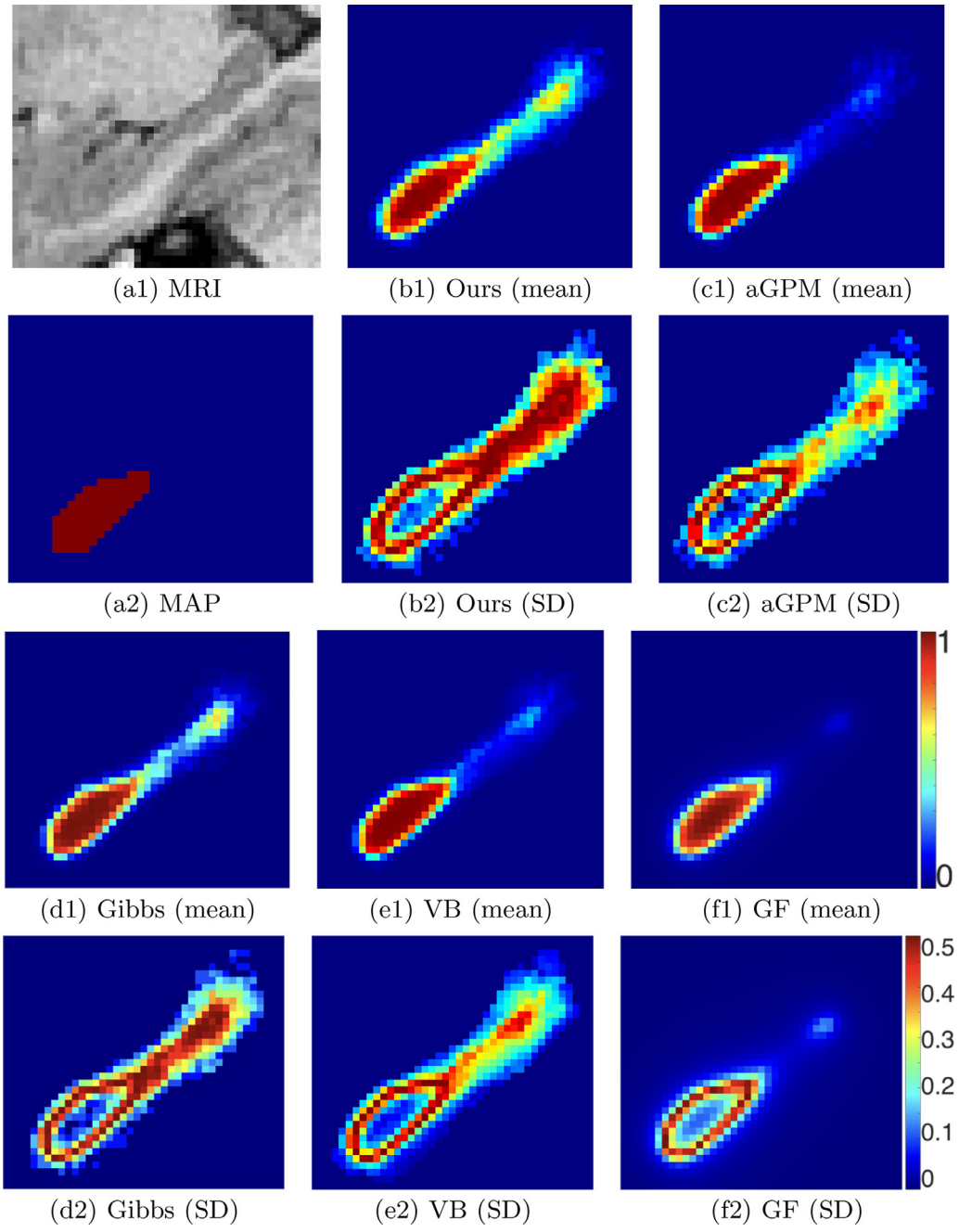

Fig. 2. Clinical brain MRI: multiatlas segmentation with ising-MRF prior, subcortical structures (Hippocampus). (a1) MRI data. (a2) MAP-MRF segmentation. (b1,b2)–(f1,f2) Voxelwise empirical label-image mean and SD (uncertainty) estimates produced from the posterior label-image distribution, using perfect sampling (our FA-BC), approximate sampling (aGPM, Gibbs), and approximate modeling (VB, GF) schemes.

5.2. Results on clinical brain MRI

We evaluate our method, and other methods, on four different kinds of applications. We use the following datasets.

1. For multiatlas segmentation of subcortical structures in brain MRI, we use data from the National Alliance for Medical Image Computing (NAMIC; http://www.na-mic.org) comprising $N = 186$ T1-weighted MRI brain images (dimensions $\approx 256 \times 256 \times 240$; voxel sizes ≈ 1 mm^3 isotropic) with expert segmentations for the caudate, putamen, thalamus, hippocampus, and globus pallidus in both hemispheres.
2. For multimodal brain-MRI tumor segmentation, we use the BRATS-17 (Menze et al., 2015) dataset.
3. For brain-MRI tissue segmentation, with a simulated lesion, we use the BrainWeb repository (http://www.bic.mni.mcgill.ca).
4. For brain-MRI parcellation, we use the NAMIC dataset.

For many segmentation tasks, MAP segmentations produced by typical methods can be very misleading by failing to expose regions with high uncertainty. The uncertainty in the segmentation of a specific part of the object can stem from the lack of information in the image data, e.g., because of low contrast, noise, and artifacts. For registration based segmentation, the uncertainty in segmentation can also arise from the difficulty in registering parts of the object reliably. These phenomena occur typically in several image segmentation problems. The hippocampus tail (Fig. 2) is a small structure with a complex shape that poses

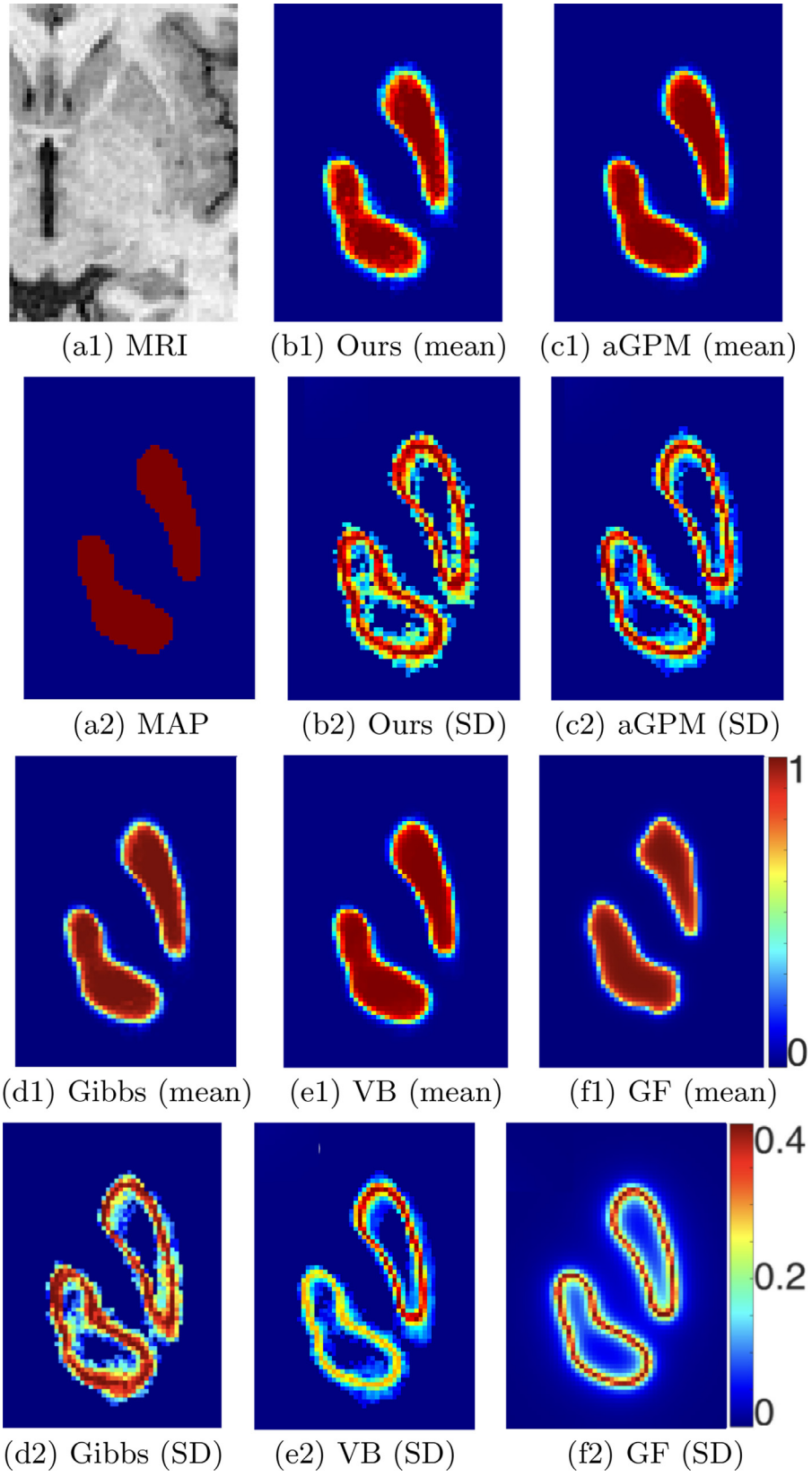

Fig. 3. Clinical brain MRI: multiatlas segmentation with ising-MRF prior, subcortical structures (Thalamus, Putamen). (a1) MRI data. (a2) MAP-MRF segmentation. (b1,b2)–(f1,f2) Voxelwise empirical label-image mean and SD (uncertainty) estimates produced from the posterior label-image distribution, using perfect sampling (our FA-BC), approximate sampling (aGPM, Gibbs), and approximate modeling (VB, GF) schemes.

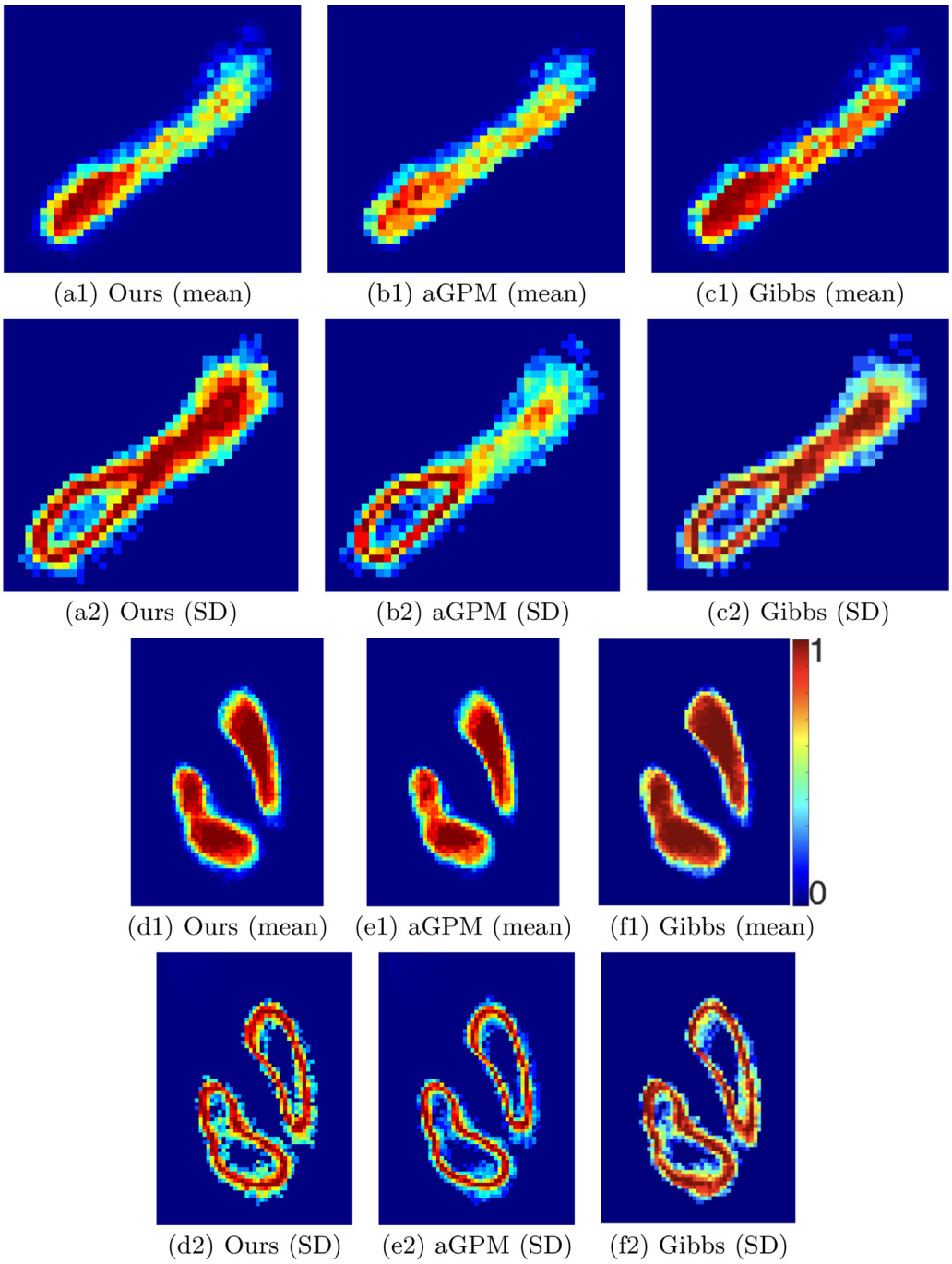

Fig. 4. Clinical brain MRI: multiatlas segmentation with SBM-MRF prior, subcortical structures (Hippocampus, Thalamus, Putamen). For the same data as in Figs. 2(a1) and 3(a1), (a1,a2)–(f1,f2) Voxelwise empirical label-image mean and SD (uncertainty) estimates produced from the posterior label-image distribution, using perfect sampling (our FA-BC) and approximate sampling (aGPM, Gibbs) schemes.

a challenge to registration methods. Several other brain subcortical structures exhibit low contrast between the intensities inside the structure and those outside, e.g., the head of the hippocampus, parts of the thalamus, and the putamen (Fig. 2). In brain tumor segmentation, the intensity contrast between the edema region and its surrounding region (Fig. 5) can be low. For brain tissue segmentation, regions with mild lesions in white matter (Fig. 6) may exhibit intensity variation leading to misclassification. In these cases, instead of the posterior mode as the output of the segmentation method, the empirical means and SDs resulting from posterior-sampled label images can be far more informative.

In multiatlas hippocampus segmentation (Fig. 2), the mean segmentation produced by our FA-BC (Fig. 2(b1)) captures the (challenging) tail portion far more accurately than all other methods (Fig. 2(c1)–(f1)), while the MAP segmentation (Fig. 2(a2)) entirely misses the tail. Our FA-BC (Fig. 2(b2)) also correctly exhibits a significant uncertainty in the tail labeling, compared to all other methods (Fig. 2(c2)–(f2)) where the low uncertainty undesirably indicates a greater confidence in a poorer result.

For multiatlas segmentation of thalamus and putamen (Fig. 3), all methods produce a reasonable labeling. However, the approximate modeling in VB and GF approaches significantly underestimates the uncertainty (Fig. 3(e2)–(f2)) compared to FA-BC.

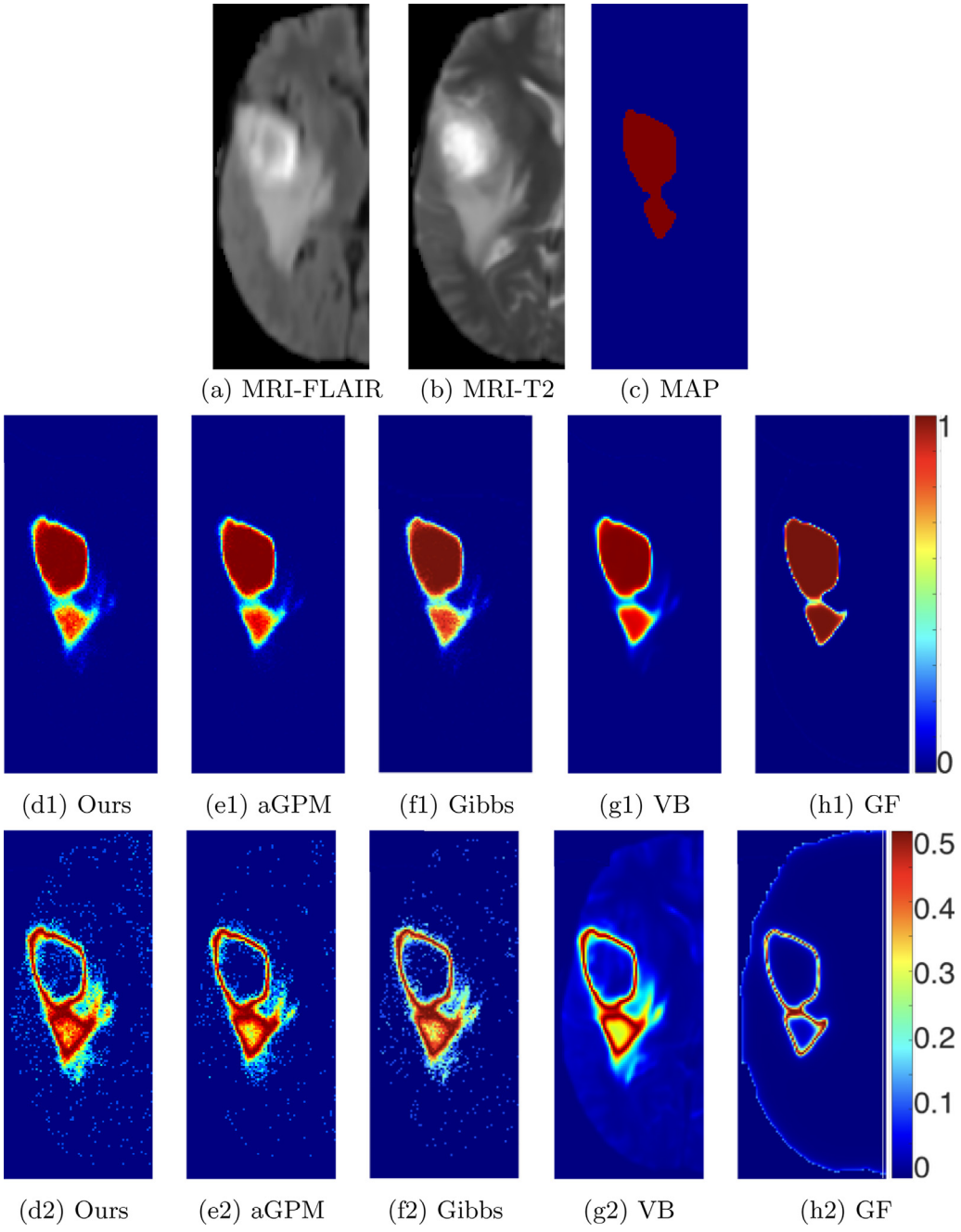

Fig. 5. Clinical multimodal brain MRI: tumor segmentation using MRF prior, GMM likelihood, and EM optimization. (a),(b) Multimodal MRI data. (c) MAP segmentation. (d1,d2)–(h1,h2) Voxelwise empirical label-image mean and SD (uncertainty) estimates produced from the posterior label-image distribution, using perfect sampling (our FA-BC), approximate sampling (aGPM, Gibbs), and approximate modeling (VB, GF) schemes.

One way to generate ground-truth uncertainty in the chosen MRI-based clinical applications could be to get manual segmentations from, say, more than twenty radiologists. Such data is typically unavailable, yet. Owing to the unavailability of ground-truth uncertainty values, we cannot quantify the performance of methods in uncertainty estimation in clinical applications. Instead, we compute Dice overlap scores between the thresholded probabilistic mean segmentation produced by each method and the ground truth. Table 1 summarizes the Dice scores (mean across subjects) for multiatlas segmentation of subcortical structures. For all structures, we see improvements from about 1% (for the putamen) to 2.5% (for the hippocampus) in average Dice score for our FA-BC compared to the next best method. Dice scores corresponding to

Table 1
Comparison of average dice scores for clinical brain MRI multiatlas segmentation of five subcortical structures with a Ising-MRF prior.

	Hippocampus	Thalamus	Putamen	Caudate	Globus pallidus
MAP	0.6720	0.8207	0.8505	0.7745	0.8064
Ours	**0.7146**	**0.8468**	**0.8701**	**0.8230**	**0.8297**
aGPM	0.6881	0.8267	0.8602	0.8086	0.8102
Gibbs	0.6885	0.8287	0.8612	0.8089	0.8116
VB	0.6811	0.8256	0.8615	0.8054	0.8087
GF	0.6769	0.8190	0.8594	0.8036	0.8093

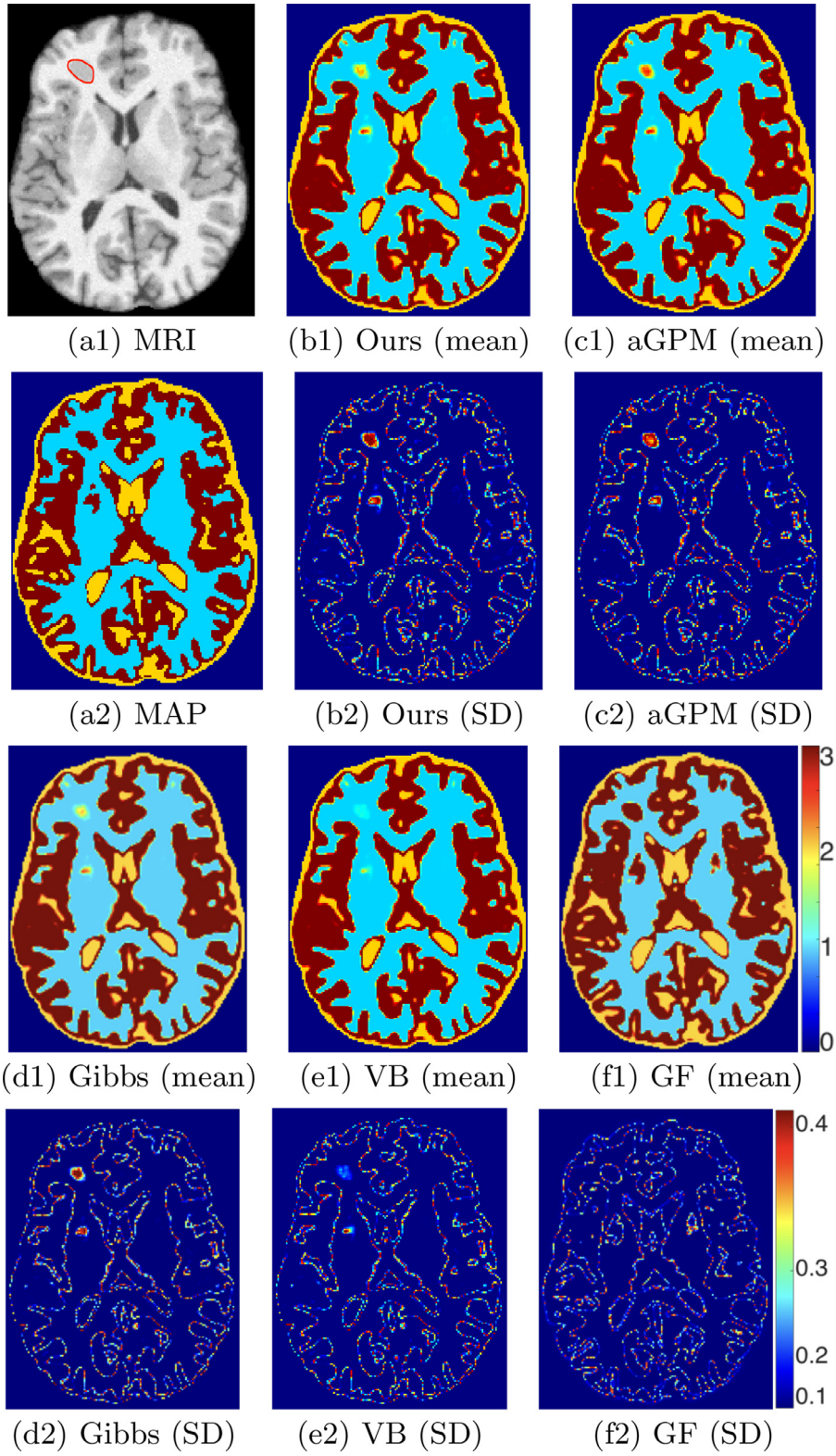

Fig. 6. Clinical brain MRI, simulated mild lesion: tissue segmentation using MRF prior, GMM likelihood, and EM optimization. (a1) MRI data with the simulated lesion outlined in red color. (a2) MAP segmentation. (b1,b2)–(f1,f2) Voxelwise empirical label-image mean and SD (uncertainty) estimates produced from the posterior label-image distribution, using perfect sampling (our FA-BC), approximate sampling (aGPM, Gibbs), and approximate modeling (VB, GF) schemes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

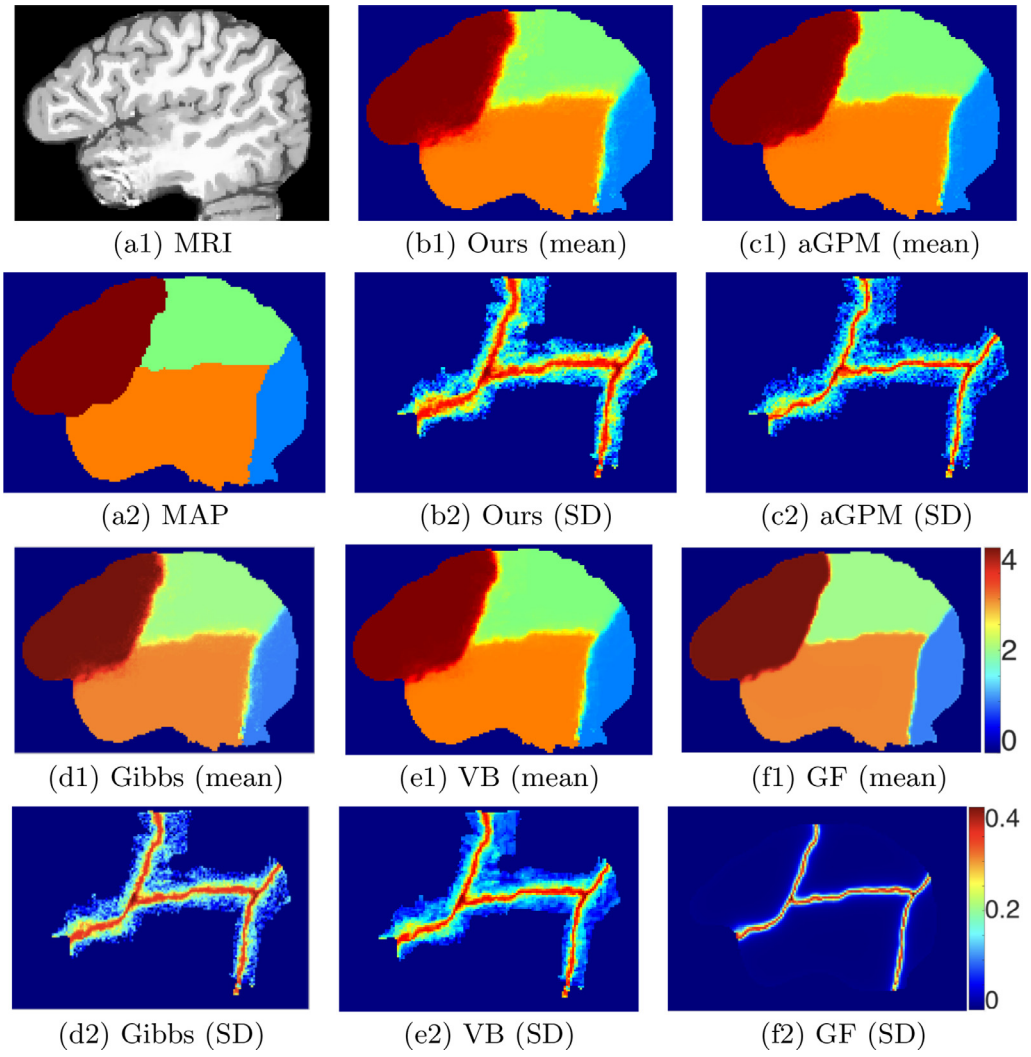

Fig. 7. Clinical brain MRI: multiatlas segmentation, Lobes. (a1) MRI data. (a2) MAP segmentation. (b1,b2)–(f1,f2) Voxelwise empirical label-image mean and SD (uncertainty) estimates produced from the posterior label-image distribution, using perfect sampling (our FA-BC), approximate sampling (aGPM, Gibbs), and approximate modeling (VB, GF) schemes.

the MAP estimates are significantly less (specifically for the hippocampus and caudate) indicating that MAP estimates alone can be highly misleading.

The SBM model (Fig. 4), compared to the MRF-models (Figs. 2 and 3), captures the tail better. This is probably because the multilayer SBM model captures the long-range correlations representing the hippocampus tail shape better than the local-neighborhood MRF model. FA-BC continues to avoid underestimating the uncertainty (Fig. 4(a2)), unlike other methods (Fig. 4(b2)–(c2)). For the thalamus and putamen, the SBM model leads to comparable results across FA-BC, aGPM, and Gibbs.

For multimodal-MRI tumor segmentation (Fig. 5), the uncertainty in the segmentation is severely underestimated by VB and GF (Fig. 5(g2)–(h2)) and moderately underestimated by aGPM (Fig. 5(e2)) and Gibbs (Fig. 5(f2)). For tissue segmentation (Fig. 6), within the mild white-matter lesion (at the top left part of image) having intensities midway between those of gray and white matter, our label mean (Fig. 6(b1)) is desirably halfway between the label values of gray and white matter. This is consistent with a greater uncertainty estimated by FA-BC (Fig. 6(b2)). While aGPM (Fig. 6(c1)) and GF (Fig. 6(f1)) label the lesion more confidently as gray matter, the VB approach (Fig. 6(e1)) labels the lesion more confidently as white matter; both these outcomes are undesirable. VB and GF significantly underestimate the uncertainty (Fig. 6(e1)-(f1)) within the lesion, while aGPM and Gibbs slightly underestimate the uncertainty. For multiatlas segmentation of lobes (Fig. 7), while GF severely underestimates the label SDs, the methods using aGPM, Gibbs, and VB underestimate SD to some extent, unlike our method that theoretically and practically guarantees sampled label images from the true posterior.

5.3. Computation-time analysis

We run all experiments on a standard workstation using an Intel Xeon processor running at 2.5 GHz (base frequency) and 64 GB RAM. Table 2 compares computation time to sample one label image from the posterior distribution, for all sampling-based methods, i.e., our FA-BC, Gibbs sampling, and aGPM. The computation times for posterior-approximation based non-sampling methods, i.e., GF and VB, are comparatively lower (about 2–4 times), because they approximate the posterior directly. Gibbs sampling convergence time varies severely with the MRF model and the data, making it very difficult to predict burn-in. The results in this paper use a typical burn-in of 5 iterations. With a safer burn-in value of about 150 iterations, where the Gibbs-sampling method gives better results, our FA-BC is roughly 8–10 times faster.

Table 2
Comparison of average time (in seconds) to obtain (i) one label-image sample for sampling-based methods (aGPM, Gibbs) and (ii) the uncertainty estimate for posterior-approximation based methods (VB, GF), for multiatlas segmentation of five subcortical structures in clinical brain MRI.

	Hippocampus	Thalamus	Putamen	Caudate	Globus pallidus
Our FA-BC	2.38	2.13	2.08	1.84	1.52
aGPM	0.41	0.41	0.40	0.38	0.31
Gibbs (burn-in 5)	0.53	0.54	0.52	0.52	0.45
Gibbs (burn-in 150)	16.82	17.36	16.31	16.08	15.10
VB	1.03	0.92	0.87	0.88	0.80
GF	0.42	0.43	0.41	0.36	0.38

6. Conclusion

We introduced a new paradigm for uncertainty estimation in segmentation relying on perfect MCMC sampling of label images from their posterior distribution. We demonstrated applicability, theoretical and practical, on (i) several MRF modeling schemes, e.g., Ising, Potts, SBM, GF, and VB, (ii) different likelihood modeling schemes, e.g., GMM and multiatlas. Our approach theoretically guarantees sampled label images to be drawn from the true posterior distribution, in finite time. This offers a significant improvement over typical approximate MCMC sampling schemes, e.g., traditional Gibbs sampling, where it is virtually impossible to predict the burn-in period, as well as the modern aGPM approach that approximates the practically intractable GPM approach. We propose to estimate uncertainty in segmentation using two perfect MCMC sampling algorithms, i.e., (i) CFTP-BC and (ii) FA-BC, where we extend the theory underlying Fill's algorithm to generic MRF models by proposing a novel BC algorithm. We show that our sampling-based approach to estimate uncertainty actually performs superior to Bayesian modeling schemes that either simplify the prior model, e.g., GF, or perform approximate variational inference, e.g., VB, even though the simplified / approximate models lead to analytical estimates of the uncertainty. We include validation using carefully designed simulated data for which we exactly know the true uncertainty, where our FA-BC outperforms all other approaches. On several classic problems in medical image analysis (segmenting tissues, subcortical structures, tumor, lobes), and several modeling and inference schemes, our results on simulated data and clinical brain MRI clearly demonstrate that our uncertainty estimates gain accuracy over several state-of-the-art inference methods.

Our approach has clear similarities to the Gibbs sampling approach. Our FA-BC is essentially tracking all possible Gibbs-sampler instances, each started from a unique state in the state space. The risk of loss in performance in the Gibbs sampler stems from improper estimation of burn-in period because of which the Gibbs sampler might return a sample that is *not* from the steady state distribution. As the burn-in for Gibbs increases, the results from the Gibbs sampler become progressively closer to the results from our FA-BC sampler. However, while estimating the burn-in of Gibbs samplers is very difficult across models, applications, and datasets, FA-BC performs that crucial task automatically and relieves that responsibility off the end-user to improve performance.

On a side note, the tracking strategy in FA-BC inflates the cardinality of the set $\mathring{X}_v$ of possible label values. This inflation in the number of tracked states in our FA-BC is analogous to the inflation in CFTP-BC that we describe in Section 3.1.3. This has the potential to reduce FA-BC's efficiency by making coalescence to occur for larger T. Nevertheless, we can view this phenomenon as sampling from the Markov chain at a later point in time.

While our method typically yields favorable results in uncertainty estimation (and segmentation), both qualitatively and quantitatively over other methods, it is computationally slower than posterior-approximation based methods (GF and VB) as well as the approximate-sampling based method of aGPM. Our method is also slower than Gibbs sampling when a very limited burn-in is used. When the Gibbs burn-in is increased conservatively to ensure sampled states to be very likely from the posterior, then our FA-BC can be significantly faster because it automatically determines termination while guaranteeing the sampled state from the stationary PMF.

Acknowledgments

The authors are grateful for support from the IIT Bombay Seed Grant (14IRCCSG010) and from the Infrastructure Facility for Advanced Research and Education in Diagnostics grant funded by Department of Biotechnology, Government of India (RD/0117-DBT0000-002).

Conflict of interest

None.

References

Alberts, E., Rempfler, M., Alber, G., Huber, T., Kirschke, J., Zimmer, C., Menze, B., 2016. Uncertainty quantification in brain tumor segmentation using CRFs and random perturbation models. In: IEEE Int. Symp. Biomed. Imag., pp. 428–431.
Awate, S., Whitaker, R., 2014. Multiatlas segmentation as nonparametric regression. IEEE Trans. Med. Imaging 33 (9), 1803–1817.
Awate, S., Zhang, H., Gee, J., 2007. A fuzzy, nonparametric segmentation framework for DTI and MRI analysis: with applications to DTI tract extraction. IEEE Trans. Med. Imaging 26 (11), 1525–1536.
Awate, S.P., Tasdizen, T., Foster, N.L., Whitaker, R.T., 2006. Adaptive markov modeling for mutual-information-based unsupervised MRI brain-tissue classification. Med. Image Anal. 10 (5), 726–739.
Awate, S.P., Tasdizen, T., Whitaker, R.T., 2006. Unsupervised texture segmentation with nonparametric neighborhood statistics. In: Proc. European Conference on Computer Vision (ECCV), 2, pp. 494–507.
Awate, S.P., Zhu, P., Whitaker, R.T., 2012. How many templates does it take for a good segmentation?: Error analysis in multiatlas segmentation as a function of database size. In: Proc. Int. Workshop Multimodal Brain Image Analysis at Int. Conf. Med. Image Comput. Comp. Assist. Interv., Lecture Notes in Computer Science, pp. 103–114.
Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. 48, 192–236.
Beutner, K.R., Prasad, G., Fletcher, E., DeCarli, C., Carmichael, O.T., 2009. Estimating uncertainty in brain region delineations. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 479–490.
Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer.
Eslami, S.M., Heess, N., Williams, C.K., Winn, J., 2014. The shape boltzmann machine: a strong model of object shape. Int. J. Comput. Vis. 107 (2), 155–1576.
Fan, A., Fisher, J., Wells, W., Levitt, J., Willsky, A., 2007. MCMC curve sampling for image segmentation. In: Med. Imag. Comput. Comp.-Assist. Interv., pp. 477–485.
Figueiredo, M., 2005. Bayesian image segmentation using Gaussian field priors. In: Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition. Springer, pp. 74–89.
Fill, J., 1998. An interruptible algorithm for perfect sampling via Markov chains. Ann. Appl. Prob. 8 (1), 131–162.
Fletcher-Heath, L.M., Hall, L.O., Goldgof, D.B., Murtagh, F.R., 2001. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. Artif.Intell.Med. 21 (1–3), 43–63.
Folgoc, L., Delingette, H., Criminisi, A., Ayache, N., 2017. Quantifying registration uncertainty with sparse bayesian modelling. IEEE Trans. Med. Imaging 36 (2).

Garg, S., Awate, S., 2018. Uncertainty estimation in segmentation with perfect mcmc sampling in Bayesian mrfs. Med. Image Comput. Comput. Assist.Interv. 21 (1), 338–346.
Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. 6 (6), 721–741.
Goh, A., Lenglet, C., Thompson, P., Vidal, R., 2011. A nonparametric riemannian framework for processing high angular resolution diffusion images and its applications to ODF-based morphometry. NeuroImage 56 (3), 1181–1201.
Gouttard, S., Styner, M., Joshi, S., Smith, R.G., Hazlett, H.C., Gerig, G., 2007. Subcortical structure segmentation using probabilistic atlas priors. In: SPIE Medical Imaging: Image Processing, 6512. International Society for Optics and Photonics, p. 65122J.
Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., Wu, X., 2011. Globally optimal tumor segmentation in PET-CT images: a graph-based co-segmentation method. In: Info. Proc. Med. Imag., pp. 245–256.
Hastings, W., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97–109.
Hazan, T., Maji, S., Jaakkola, T., 2013. On sampling from the Gibbs distribution with random maximum a-posteriori perturbations. In: Neural Info. Proc. Sys., pp. 1268–1276.
Huber, M., 2004. Perfect sampling using bounding chains. Ann. Appl. Prob. 14 (2).
Iglesias, J., Sabuncu, M., 2015. Multi-atlas segmentation of biomedical images: a survey. Med. Imaging Anal. 24 (1), 205–219.
Jena, R., Awate, S.P., 2019. A Bayesian neural net to segment images with uncertainty estimates and good calibration. In: Info. Prod. Med. Imag., pp. 1–13.
Kader, G., Perry, M., 2007. Variability for categorical variables. J. Stat. Educ. 15 (2), 1–17.
Kohli, P., Torr, P., 2008. Measuring uncertainty in graph cut solutions. Comput. Vis. Imaging Underst. 112, 30–38.
Koller, D., Friedman, N., 2009. Probabilistic Graphical Models: Principles and Techniques. MIT press.
Kybic, J., 2010. Bootstrap resampling for image registration uncertainty estimation without ground truth. IEEE Trans. Imaging Proc. 19 (1), 64–73.
Le, M., Unkelbach, J., Ayache, N., Delingette, H., 2016. Sampling image segmentations for uncertainty quantification. Med. Imaging Anal. 34, 42–51.
Li, S.Z., 2009. Markov Random Field Modeling in Image Analysis. Springer.
Menze, B., Jakab, A., et al., S.B., 2015. The multimodal brain tumor image segmentation benchmark BRATS. IEEE Trans. Med. Imaging.
Papandreou, G., Yuille, A., 2011. Perturb-and-MAP random fields: using discrete optimization to learn and sample from energy models. In: Int. Conf. Comp. Vis, pp. 193–200.
Perry, M., Kader, G., 2005. Variation as unalikeability. Teach. Stat. 27 (2), 58–60.
Pham, D., Prince, J., 1999. A generalized EM algorithm for robust segmentation of magnetic resonance images. In: Info. Sci. and Sys., pp. 558–563.
Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M., 2007. A hierarchical algorithm for MR brain image parcellation. IEEE Trans. Med. Imaging 26 (9), 1201–1212.
Propp, J., Wilson, D., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms 9 (1), 223–252.
Sabuncu, M.R., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P., 2009. Supervised nonparametric image parcellation. In: Medical Image Computing and Computer-Assisted Intervention, pp. 1075–1083.
Salakhutdinov, R., Larochelle, H., 2010. Efficient learning of deep Boltzmann machines. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 693–700.
Salakhutdinov, R., Mnih, A., Hinton, G., 2007. Restricted Boltzmann machines for collaborative filtering. In: Proceedings of the 24th International Conference on Machine Learning, pp. 791–798.
Shah, M., Bhalgat, Y., Awate, S.P., 2019. Annotation-cost minimization for medical image segmentation using suggestive mixed supervision fully convolutional networks. In: Med. Imag. Meets Neural Info. Proc. Sys., pp. 1–4.
Shah, M., Merchant, S.N., Awate:, S.P., 2018. MS-Net: mixed-supervision fully-convolutional networks for full-resolution segmentation. In: Med. Imag. Comput. Comput.-Assist. Interv., pp. 379–387.
Song, Z., Awate, S.P., Licht, D., Gee, J., 2007. Clinical neonatal brain MRI segmentation using adaptive nonparametric data models and intensity-based Markov priors. In: Proc. Med. Image Comput. Comp. Assist. Interv., 1, pp. 883–890.
Song, Z., Tustison, N., Avants, B., Gee, J.C., 2006. Integrated graph cuts for brain MRI segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 831–838.
Veni, G., Fu, Z., Awate, S.P., Whitaker:, R.T., 2013. Bayesian segmentation of atrium wall using globally-optimal graph cuts on 3D meshes. In: Info. Prod. Med. Imag., pp. 656–667.
Williams, C., Barber, D., 1998. Bayesian classification with Gaussian processes. IEEE Trans. Pattern Anal. Mach.Intell. 20 (12), 1342–1351.
Wolz, R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., 2009. Segmentation of subcortical structures and the hippocampus in brain MRI using graph–cuts and subject-specific a-priori information.. In: IEEE Symp. Biomed. Imag., pp. 470–473.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden markov random field model and the expectation maximization algorithm. IEEE Trans. Med. Imaging 20, 45–57.